8-10-2021

Characterization of the Contributions of Hydration in Protein-DNA Specificity via Structurally Conserved ETS Family Transcription Factors PU.1 and Ets-1

Amanda V. Albrecht

CHARACTERIZATION OF THE CONTRIBUTIONS OF HYDRATION IN PROTEIN-DNA SPECIFICITY VIA STRUCTURALLY CONSERVED ETS FAMILY TRANSCRIPTION FACTORS PU.1 AND ETS-1

by

AMANDA V. ALBRECHT

Under the Direction of Gregory M. K. Poon, PhD

ABSTRACT

Protein-DNA interactions involve a redistribution of conformational dynamics and hydration properties. As a model DNA-binding protein, structural and thermodynamic studies have shown that ETS transcription factor PU.1 DNA selection is dependent on solvent reorganization. PU.1 high-affinity DNA binding is osmotically sensitive, whereas low-affinity binding is not, which provides a model for examining the role of hydration and DNA dynamics in protein-DNA specificity. Characteristic of ETS members, PU.1 recognizes DNA sequences harboring a central 5'-$G_0G_1A_2A_3$-3' consensus, where an Arg sidechain contacts O^6 of G_1 in the major groove. This conserved interaction presented an opportunity to isolate the contributions of

DNA conformational dynamics and interfacial hydration from direct readout via chemically defined deoxyguanosine derivatives. As a control, the substitution of G_1 with 2-aminopurine (d2-AP) eliminates O^6 in the major groove, resulting in low-affinity binding. Exchanging O^6 with N^2H_2 in G_1 (iso-dG/iso-d^{5m}C) produced the same result, suggesting that C=O is an irreplaceable H-bond acceptor in ETS/DNA direct readout. Removal of NH_2 (inosine) in the minor groove, opposite the protein-contact surface, impaired binding without altering the osmotic properties. Inosine substitution weakened local base-pairing, which suggests dynamics and hydration contributions in PU.1/DNA recognition could be separated. To explore this possibility, we assessed the binding of the high-affinity 5'-GIAA-3' sequence to chimeric PU.1, where the DNA-contacting surfaces were partially interchanged with those from Ets-1, an ETS-family relative which is insensitive to osmotic pressure. The osmotically less sensitive chimeras were also less susceptible to inosine substitution. The modified nucleobases show that dynamic and hydration contributions (inosine, dI) may be separated from direct readout contributions (d2-AP, Iso-dG). The differential sensitivity of PU.1/Ets-1 chimeras to inosine-substituted DNA suggested an evolution of PU.1 from ETS relatives in which excess hydration is coupled to restraining DNA dynamics in complex formation.

INDEX WORDS: Specificity, Transcription factor, ETS family, PU.1, Ets-1, Hydration, Thermodynamics, DNA

CHARACTERIZATION OF THE CONTRIBUTIONS OF HYDRATION IN PROTEIN-DNA SPECIFICITY VIA STRUCTURALLY CONSERVED ETS FAMILY TRANSCRIPTION FACTORS PU.1 AND ETS-1

by

AMANDA V. ALBRECHT

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

CHARACTERIZATION OF THE CONTRIBUTIONS OF HYDRATION IN PROTEIN-DNA SPECIFICITY VIA STRUCTURALLY CONSERVED ETS FAMILY TRANSCRIPTION FACTORS PU.1 AND ETS-1

by

AMANDA V. ALBRECHT

Committee Chair: Gregory M. K. Poon

Committee: W. David Wilson

Markus W. Germann

Giovanni Gadda

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2021

DEDICATION

I dedicate this dissertation to my mom, Joy Albrecht, for all her love and support.

In memory of my grandparents Gary and Virginia Albrecht.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Gregory Poon, my advisor. I want to thank him for giving me this position in his lab and for providing an environment in which to grow and learn as a scientist, for all his guidance, both academically and scientifically. I am grateful for his constant drive to be better, to learn, and to present our findings before coworkers and at conferences. With his support, I had the opportunity to mentor other graduate students and undergraduates, which helped expand my skills as a leader and an educator. With his direct mentorship, I have learned many new skills and techniques, and I appreciate the time he took to personally teach me and help me become a better scientist. I certainly would not be here today without his guidance and encouragement.

To my committee, Dr. David Wilson, Dr. Markus Germann, and Dr. Giovanni Gadda, I would like to thank them for all their help and input throughout this process. Thank you specifically to Dr. Wilson for his guidance throughout these past years. Thank you for the insightful questions in lab meetings and for providing me the opportunity to practice and learn how to present scientific data in front of my peers. I would also like to thank Dr. Germann for his patience and in helping me learn NMR techniques through his class and data collection. Thank you for your guidance and encouragement. To Dr. Gadda, I would like to thank him for the education I have gained through his class and his feedback on this project. I want to thank him for his insights and unique way of looking at the data and for helping to build my confidence. It has been a pleasure to work with him as a scientist and as a part of the chemistry graduate student association.

None of this would be possible without the help and support of my lab mates, both past and present. I would like to gratefully thank Dr. Suela Xhani, Dr. Sangchoon Lee, Dr. Shingo

Esaki, Van Ha, Noa Erlitzki, Kenneth Huang, and Hye Mi Kim. Specifically, I would like to thank Suela and Ken for their constant encouragement and support throughout this journey. We started this process and have grown together through all the ups and downs; I would not be where I am today without the two of you, and for that, I am eternally grateful. I will cherish all of the memories, the lively discussions, and the inside jokes. You made the late nights and long days enjoyable, even when it meant just turning up the music. Thank you for pushing me to be better and work harder and, of course, for all the coffee breaks. Thank you for being amazing friends and coworkers.

I want to acknowledge my inspiring first PI, Dr. Anita Corbett. Without her support, I would not be in graduate school and finishing my Ph.D. I cannot thank her enough for taking a chance on me and for allowing me to learn what it means to work in a lab and what goes into scientific research. She gave me an internship in her lab without having any research background, gave me my first project, and inspired me to become the scientist I am today. Thank you, Anita, for being an inspiration to so many up-and-coming scientists and for giving me this opportunity.

Thank you to my dear friend Dr. Annie McPherson, my first graduate student mentor in Dr. Corbett's lab. Thank you for your patience and guidance as I learned how to be a research scientist. Without Annie, I would not have the strong scientific foundation that I have since built upon. Thank you for getting me started on this path and for being a fantastic friend throughout this period in my life, for her support and encouragement through everything life and graduate school has thrown my way.

They say it takes a village to achieve your dreams, and that cannot be truer. My village is my family; I would like to thank my aunts, uncles, and cousins for all of their support and encouragement through this time.

I cannot go any further without acknowledging and thanking my amazing mother, Joy Albrecht. Her never-failing confidence that I would succeed was a lifeline during the most challenging moments. Throughout this adventure, knowing that she was always a phone call away, a shoulder to cry on, my rock and constant companion on this journey was the biggest blessing and comfort. Thank you for her guidance and encouragement to overcome any obstacles in the way. I would never have gotten here today without my mom as my role model; she has always been a pillar of strength and support whenever I needed her. Thank you for teaching me how to persevere and for giving me your determination. Thank you for raising me to reach for my dreams, to keep pushing for what I want, for giving me a never-ending love of learning. Thank you for all that you have sacrificed, for leading by example, and for showing what hard work can do. Thank you for being my role model and someone I will forever admire and look up to.

My grandparents, Gary and Virginia Albrecht, were some of my biggest supporters and cheerleaders; it is hard to believe that they are not here to see this day after all that they have done for me. I would not be where I am today if it were not for them. Thank you for your guidance and support throughout my life and this journey. Thank you for being outstanding examples of how to raise and support your family, how to put in the hard work, and share with those around you. I am eternally grateful for all you have done for me and for always being there.

To my brother Robert and sister Annie, thank you for always making me laugh and for your never-ending support. Your prayers and encouragement mean the world to me, and I certainly would not be who I am today without the two of you.

Finally, to my amazing husband, Chris Reed, thank you for standing by me through all of this. Thank you for your never-ending patience as life became stressful and chaotic. You have been a fantastic support to me, from our first date the weekend before my qualifying exams, to our wedding, to defending in a pandemic, and as we move forward to welcome our beautiful son into this world. Thank you for being a shoulder to cry on when I did not think I could keep going and for encouraging me through my failures and successes. I cherish our evening walks when we can decompress from the day and discuss any issues or successes that arose and brainstorm together how to move forward. Your joy and optimism are infectious; you helped me believe in myself and keep pushing forward to my goal. Thank you for all that you do and for being such an amazing husband and the perfect person to experience life with as we move forward on this next great adventure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Deoxyribonucleic acid (DNA)
Transcription factor (TF)
DNA binding domain (DBD)
E26 transformation-specific (ETS)
Wildtype (WT)
Integration Host Factor (IHF)
Nuclear Magnetic Resonance (NMR)
Nuclear Overhauser effects (NOE)
Nuclear Overhauser effect spectroscopy (NOESY)
Rotating frame nuclear Overhauser effect (ROE)
Rotating frame nuclear Overhauser effect spectroscopy (ROESY)
Pointed domain (PNT)
On-SighT domain (OST)
Isothermal Titration Calorimetry (ITC)
Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq)
Guanosine (dG)
Adensoine (dA)
Cytosine (dC)
Thymine (dT)
Inosine (dI)
2-Aminopurine (d2-AP)
Iso-Deoxyguanosine (Iso-dG)
Iso-Deoxycytosine (Iso-dC)
Iso-Deoxy-5-methylated-cytosine (Iso-d^{5m}C)
Molecular dynamics (MD)

1 GENERAL INTRODUCTION

1.1 The importance of affinity and specificity in protein-DNA interactions

The study of protein-DNA specificity is critical for a clear understanding of biological processes such as transcription, gene regulation, DNA repair, and additional applications, including drug design [1-4]. Through the knowledge of how proteins and DNA bind, and the mechanisms that drive these interactions, it is possible to understand how the cell differentiates between transcription factors (TF) that utilize the same overall structure and bind to the identical cognate sequences, all the while these TF can signal for differing effects to occur. The underpinnings of specificity have long been studied, as it is key to the proper functioning of the cell in all living organisms; many proteins within the cell require specific binding to carry out their functions. One such class of proteins are restriction enzymes which act to cleave DNA at key locations [1, 5]. By understanding the specificity of the interaction, it is possible to harness the specific functions for scientific advances- the sequence specificity of restriction endonucleases has been utilized in molecular cloning and the creation of recombinant DNA. Restriction endonucleases work to cut DNA in specific locations, palindromic sites, and can create an overhang on both ends of DNA, which provides a method for easily inserting or removing lengths of DNA from a sequence [6, 7]. Another goal of understanding affinity and specificity is in the development of therapeutics and drug design. Understanding what is required to bind with a specific site on DNA provides a greater ability to design and synthesize drugs to target specific sites on DNA. This growing knowledge provides greater avenues for therapeutics to block protein-DNA interaction or enhance these interactions. Achieving specificity requires multiple factors, including the DNA sequence, structure of both DNA and protein, and whether the protein interacts directly or indirectly with the DNA [1, 8, 9].

Protein-DNA interaction occurs with varying levels of affinity, which helps to explain specificity. Affinity is the strength of which one molecule binds to another; in the case of proteins binding to DNA affinity is measured in terms of the association (K_a) or dissociation constant (K_d) (equation 1-1) [10].

$$K_d = \frac{[\text{Protein}][\text{DNA}]}{[\text{Protein:DNA}]} \qquad \textit{Equation 1-1}$$

Affinity is obtained through various noncovalent interactions, including hydrogen bonds, van der Waals, electrostatics, water-mediated, and hydrophobic interactions; these help decrease the overall binding free energy allowing the interaction to take place [3, 10, 11]. In the case of binding affinity, the free energy of interest is the difference between the free energy of the unbound structure vs. the free energy of the bound structure [10]. Whereas specificity is defined as the difference in affinity of one complex vs. another complex [10]. In many cases, the level of affinity and specificity are typically coupled, where a protein that binds DNA with high affinity will also display higher levels of specificity. However, when dynamics and structure are brought into the picture, this coupling may change. A protein binding to a flexible DNA, or a flexible protein binding to DNA may have a lower binding affinity while still retaining high levels of specificity [10].

Therefore, the structures of the proteins and DNA must also be taken into account when examining affinity, as contact is required to achieve any level of specificity- factors that affect the level of contact between the protein/DNA include counterions and levels of interfacial hydration across the surfaces of proteins and DNA in the bound and unbound states [3, 11]. The stronger the affinity of the protein for a specific sequence tends to correspond to greater

specificity; in other words, a protein that binds weakly with a variety of sequences will have lower specificity, while a protein that binds strongly to a few select sequences will have greater specificity [4, 12]. For a protein and DNA to achieve the strongest affinity, the shape of the binding sites must be complementary, along with an equally compatible charge distribution across the binding interfaces [4]. The shape complementarity of the protein-DNA interactions drives indirect readout. In contrast, the specific consensus sequence drives direct readout, both of which are required to achieve high-affinity and high specificity for protein-DNA interactions and will be discussed in greater detail in later sections.

Specificity is a fascinating aspect of biological interactions; previous research has shown that oligonucleotides can differentiate between different isomers of the same amino acid and between cofactors that are oxidized vs. reduced [4]. This leads to the question of how specificity is determined? We know that specificity is affected by the level of binding affinity of the complex and that minute changes in the complex can lead to drastic alterations in specificity, but how do we explore specificity? Equilibrium thermodynamics are essential for understanding protein-DNA specificity, where specificity is determined by the differences in binding free energy between specific and non-specific interactions. Binding free energy itself is shown by the Gibbs free energy equation (equation 1-2) [3].

$$\Delta G = \Delta \mathrm{H} - \mathrm{T}\Delta S \qquad \textit{Equation 1-2}$$

The affinity of protein-DNA interactions is achieved through some level of direct interaction between the interfaces- at the interfacial surface, the protein, DNA, and/or hydration layer may change to increase/decrease affinity. These changes are measurable through the change in the heat of the reaction (ΔH), the change in hydration (Δn_w), and the change in volume (ΔV). By

either calculating the elements of the Gibbs free energy directly or altering the environment of the interactions, it is possible to glean insight into what drives protein-DNA specificity. Through rearranging the Gibbs free energy equation, it is possible to determine the change in free energy (ΔG) through changes in binding affinity (equation 1-3):

$$\Delta G^{O} = RT \ln K_d \qquad \textit{Equation 1-3}$$

Generally, high-affinity binding events lead to high specificity in protein-DNA interactions; however, there are cases where the protein binds DNA at low-affinity levels but still retains the ability to differentiate against other DNA sequences leading to higher levels of specificity than previously expected [13].

Many events affect equilibrium thermodynamics, such as the release of water from the surface of the protein and DNA upon complex formation and conformational changes in protein backbone and sidechain to interact with DNA effectively [3]. By examining the effects of changing the conditions of protein-DNA interactions, it is possible to calculate the thermodynamic costs and better understand what is required for protein-DNA specificity.

1.1.1 Direct Readout

Direct readout refers to when the residues of the protein and the DNA bases interact through hydrogen bonds along the DNA grooves, whether directly, via water-mediated hydrogen bonds, or through ionic interactions between complementary charges [3, 8, 9, 14, 15]. The DNA sequence is critical for proper specific interactions; with the appropriate sequence, the protein can bind into either the major or minor groove of DNA [1]. Where alpha helices are inserted into the major groove of DNA to obtain direct readout through van der Waals interactions, in a few cases, a beta-sheet may be inserted instead of an alpha helix [11]. The insertion of an alpha helix

into the major groove of DNA with the accompanying van der Waals interactions generally leads to a decrease in available surface area, leading to the exclusion of water from the binding site [11].

Many proteins bind DNA through specific sequences called a consensus sequence, and the strength of the binding affinity depends strongly on the accuracy of this consensus sequence [16]. Any changes in the consensus sequence may lead to decreased binding affinity; screening potential DNA sequences makes it possible to determine if a protein will interact with any sequence based on whether the consensus site is present. This is a primary example of the role that direct readout of DNA plays in allowing the protein to achieve sequence specificity [16].

1.1.2 Indirect Readout

Protein-DNA interactions are critical for achieving specificity, but direct readout is only one method of acquiring the proper interactions. Another method for attaining specificity is indirect readout. Indirect readout combines with direct readout to achieve specificity and strong binding affinity when bound to cognate DNA sequences [14]. Indirect readout helps the protein form partial specificity with a sequence and occurs through DNA and protein structure recognition [17]. Indirect readout considers how each shape changes upon interaction- either by the DNA, protein, or other binding partners and, importantly, the DNA's flexibility [3, 9, 18]. The conformations involved in indirect readout are linked to structural properties such as shape and width of the DNA major and minor grooves, the DNA curvature, the level of hydration in and around the protein or DNA, the flexibility of the DNA, and the DNA backbone conformation [15, 19]. In contrast, direct readout is dependent on precise contacts formed by nucleic acids and amino acids, whereas indirect readout considers bases not directly interacting with the protein. Indirect readout also factors in any water-mediated contacts, small molecule interactions, and the

conformation of the sugar-phosphate backbone of DNA [14, 15, 17, 19]. The complementary nature of DNA means changes in the conformation of bases outside the binding interface will influence the affinity of the interaction by affecting bases in direct contact [14]. Unbound protein and DNA take on various conformations dependent on the environment, which may differ from the structures achieved when the complex is formed. When forming a complex, the binding site must be made available to the binding partner, potentially changing the overall structures. The DNA and protein structures may vary significantly depending on the flexibility of the structure, the environment, and the number and type of binding partners [14]. The conformational change requires energy due to the forming and breaking of bonds. This energy cost also plays a role in determining the complex's affinity and specificity, as demonstrated in the Gibb's free energy equation (equation 1-2) [14].

One form of indirect readout is that upon protein binding DNA, the DNA conformation changes through either cooperative or anti-cooperative base coupling. This allows the DNA to adopt the most stable conformation to provide the complex with better specificity [20, 21]. However, unbound DNA free in solution with a conformation already similar to what the protein requires to bind will have a stronger binding affinity than a DNA sequence that must undergo a drastic conformational change to enable binding [14]. The higher energetic cost required for a considerable conformational change before binding negatively impacts the affinity of the final complex. Likewise, more flexible DNA that can adopt multiple shapes will bind more targets with lower binding affinity due to the energy cost of changing shape, therefore having lower specificity. A stiffer DNA in the proper shape that compliments the desired binding partner will bind with stronger affinity and higher specificity but will be unable to bind as many partners [4, 14].

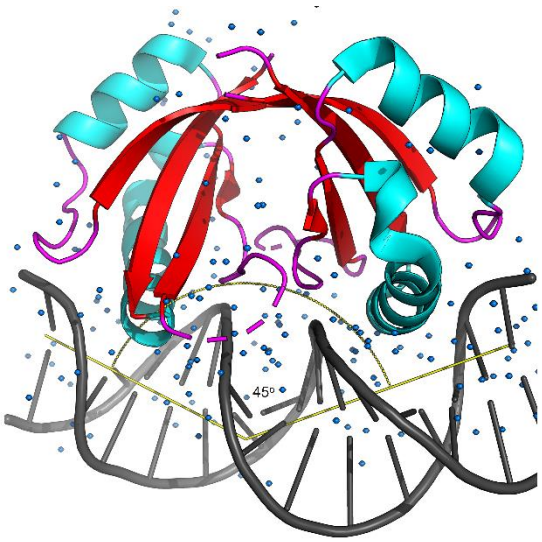


Figure 1-1 Human papillomavirus (HPV) E2 bends DNA upon binding to form a specific complex

Upon binding DNA, HPV E2 (1JJ4, [22]) bends the DNA by 45° towards the minor groove where the four spacer bases are located [23]. The bend of the DNA around the spacer nucleobases happens regardless of the identity of the spacer bases, but their identity does alter binding affinity even though there are no direct interactions between the protein and the spacer bases [23].

DNA bending to fit the proper conformation with the protein has been well examined through crystal structures. In the catabolite gene activator protein crystal structure with specific DNA, the DNA is bent about 90° [11]. The cyclic AMP receptor protein (CRP) is another example of a protein that binds DNA selectively through direct and indirect readout by bending the DNA by 90° upon binding [19]. CRP binds DNA through a helix-turn-helix motif making direct interactions with multiple DNA bases. In the crystal structure, the bend is caused by a primary and secondary kink in the DNA [19]. The bend in the CRP/DNA complex is supported through electrostatic interactions between the protein and the phosphate backbone of DNA [19]. What makes indirect readout more attractive for some protein/DNA complexes is the allowance for variation in the protein or DNA sequence. In CRP/DNA complexes, the DNA consensus sequence is retained; however, the genome of different CRP proteins is divergent, suggesting that the primary sequence does not drive high-affinity interactions. Instead, the bending of DNA is the principal factor for induced fit interactions [19]. The human papillomavirus (HPV) E2 bends DNA by 45° upon binding; however, it does not form any direct hydrogen bonds between

the protein and the spacer units on the DNA, even though previous work has proven that the sequence of the spacer unit is variable and affects the level of binding affinity of the complex [23]. Variations in the spacer sequences of the DNA that binds E2 may work to either attract or repel binding partners through various distributions of charges across the binding interface, which are believed to play a role in the long-range electrostatic interactions [23].

Bending is a pervasive structural alteration for the DNA to obtain efficient interactions with binding partners. However, bending is not the only structural change that DNA can make to achieve specificity. In the integration host factor (IHF), the DNA twists to achieve indirect readout [11]. IHF is a classic example of a protein that binds DNA through indirect readout. The complex only makes three direct interactions with DNA but retains high-specificity interactions despite the lack of contacts [15].

The environment in which the DNA resides also plays a vital role in the overall structure- if the solution has a high concentration of cations, the shape will adjust to fit. These alterations can be seen in the groove width, along with increased or decreased bending and twisting [14]. Many monovalent cations have been found in the minor groove floor, resulting in a narrowing of the minor groove in AT-rich sequences [14]. The narrowing of the minor groove is thought to be a method of indirect readout, making the major groove a more favorable binding site for proteins by adjusting the shape of the DNA grooves to accommodate the protein binding site [14]. On the other hand, divalent cations are larger and are more likely to be found in the major groove of DNA, which leads to the DNA bending rather than the opening or narrowing of the grooves [14]. Additionally, intrinsic forces such as hydrophobicity and London dispersion forces cause the DNA bases to stack atop one another automatically. These weak forces act to restrain the DNA structure in its native form [14]. All of these changes to the protein or DNA shape, along with the

addition or removal of any water-mediated contacts or small molecules, provides the protein a means of differentiating between sequences energetically [17].

Changes to the internal amino acids of proteins also change protein-DNA binding affinity, even when the DNA sequence is identical. This demonstrates that even though the binding site is identical, and the protein and DNA make the same direct interactions, it is possible to achieve a low-affinity complex due to indirect readout changes, as seen in the lambdoid bacteriophage repressor-DNA complex [14].

1.1.3 Hydration

Water molecules on the surfaces of proteins and DNA tend to exchange rapidly with bulk solution. However, water-mediated interactions important for protein-DNA and protein-protein binding tend to have longer lifetimes than bulk solution [24]. Water helps the protein in a multitude of structural aspects- the hydrophobic effect promotes the nonpolar amino acid residues to congregate away from water, which leads to protein folding. Once correctly folded, the protein will interact with DNA, and waters will generally be released from the interfacial surface [11]. In protein structures, the proteins that form a ‘drier’ interface have fewer cavities. In contrast, a protein with a more hydrated interface will generally have various cavities filled with water [24]. Water also fills in solvent excluding cavities within the tertiary protein and plays an essential role in polar interactions providing more structural stability [3, 24]. Protein cavities generally have one to three water molecules within the cavity, and these waters are stabilized through hydrogen bonds and are highly ordered [25]. This ordering of water in cavities away from bulk solution results in an increase in enthalpy and decrease in entropy, with an overall favorable change in free energy [25]. Water plays varying roles thermodynamically on proteins depending on the protein's flexibility- a more rigid protein will result in an entropy penalty to

hold onto the waters while gaining enthalpy [26]. Likewise, a flexible protein releasing waters in bulk solution serves to increase entropy while having a decrease in enthalpy [26].

In DNA, water is essential for the proper formation of B-form DNA. At AT-rich sites, water is commonly found in the minor groove and is thought to help stabilize the structure [11, 27]. In the Drew and Dickerson B-DNA sequence, the minor groove was discovered to contain a spine of hydration consisting of ordered waters that make up two to four layers of hydration [28]. The hydration shell is made up of waters directly around the DNA; in B-form DNA, the hydration shell interacts with the oxygens on the phosphate backbone. In A and Z-form DNAs, the phosphates are closer together, and the water molecules act as bridging contacts between the phosphate backbone [27]. To explore the residency times of water within the hydration shell of DNA, one can measure nuclear Overhauser effects (NOEs) and rotating frame nuclear Overhauser effects (ROEs) to find the cross peaks [27, 29]. When examining DNA hydration in the rotating frame nuclear Overhauser effect spectroscopy (ROESY) spectrum, at the frequency of water, DNA protons near water will display a negative cross peak to indicate exchange with water. However, in the nuclear Overhauser effect spectroscopy (NOESY) spectrum, any DNA proton interacting with water will show a positive cross peak if the residency time is more than 500 ps; anything shorter (<500 ps) will result in a negative cross peak [27, 29, 30]. Therefore, NOEs can be used to directly estimate residency time of waters along the surface of macromolecules like DNA or proteins [29, 30]. Waters found around DNA are vital, and through studying the levels of hydration around DNA, it is possible to predict the locations of protein-DNA binding [28, 31]. The level of hydration of protein binding sites on DNA is highly dependent on the DNA sequence and conformation. To examine this, a series of unbound DNA crystal structures were examined to determine the level of hydration before and after complex

formation, with close attention paid to conserved DNA sequences [31]. The study found that 86% of the DNA binding sites were distinguished by the hydration of the complexes examined. In the *trp*-repressor-operator complex, 24 of the 25 water-mediated contacts were successfully predicted by hydration sites to within 1.5Å [31].

In DNA, it is more common to see water interacting with the backbone rather than the bases or sugar atoms; in proteins, glutamine (Glu) and aspartate (Asp) tend to make the most contact with water [11, 24]. Even though DNA tends to be more polar due to the phosphate backbone and tends to have an overall negative charge, DNA-protein complexes can be either hydrated or dehydrated [24]. Water-mediated contacts between protein and DNA appear to have longer residence times than previously believed- when examined by nuclear magnetic resonance (NMR), it is possible to measure the NOEs of the water-mediated contacts between the protein and DNA of interest. The NOEs of the water molecules on the protein-DNA interface may last in the nanosecond (10^{-9}) range, and waters secluded within the protein can have residency times as far as the upper milliseconds (10^{-4}) [24].

The level of hydration and flexibility of the DNA is very much dependent on the sequence and plays an important role in protein-DNA interactions [28]. DNA sequences that are the most structurally constrained are more likely to have ordered hydration patterns, whereas more flexible DNA sequences are less likely to have any ordered hydration [28]. The rigidity of DNA depends on sequence composition- sequences containing higher levels of AT base pairs tend to be more rigid and more hydrated, whereas swapping the bases to TA results in a more flexible stretch of DNA [28]. During the formation of protein-DNA complexes, the waters around the interaction surface may be excluded and released into bulk solvent as they are displaced. This results in the dehydration of the interaction surface, as displayed by the hydrophobic GATA.1-

DNA complex [3, 11, 28, 32, 33]. When these waters of the protein-DNA interface are released, there is an increase in the system entropy, helping to promote the interactions between the protein and DNA [28]. The release or retention of hydration in the protein-DNA interface can play a significant role in binding affinity due to the changes in entropy and enthalpy [28]. A large release of water upon the AT-hook protein binding in the minor groove of DNA results in a large increase in entropy, thus driving the binding interactions [28].

Previously it was thought that when a protein binds to DNA, the interactions created replace water-mediated interactions of the unbound components due to the reduction of solvent accessible surface area [11, 34]. This release of water upon protein-DNA binding causes an increase in the system's entropy, driving binding thermodynamically [34]. Removal of water at the interface is not the only method of changing the level of hydration of the system; allosteric changes may be triggered upon protein-DNA binding that can alter the structure of the protein or DNA at areas distal to the binding site, changing the surface area available to solvent [34]. However, this is not always the case- the *trp* repressor-operator complex is a well-known example of a highly hydrated sequence-specific complex. In the *trp* repressor-operator complex, three bases in the operator site are well hydrated and may act as a recognition signal [3].

The changes in hydration have been examined in many proteins; in these experiments, water is expected to be excluded from the binding surface as seen when hemoglobin binds oxygen and when restriction endonuclease *Eco*RI binds high-affinity DNA [33, 35-37]. This dehydrating effect is seen in crystal structures of specific complexes; when crystal structures of unbound DNA and protein are compared with the bound complex, the bonds to water usually are replaced with bonds between the protein and DNA [38]. However, sometimes water is sequestered upon binding, as with *Eco*RI to nonspecific DNA, which retains ~110 more waters

than in the high-affinity complex. These waters are expected to be found in cavities that exclude solutes [33, 35, 38-40]. The increased water concentration of the nonspecific complex will most likely be found in the protein-DNA interface since the level of sequestered water is not affected by osmolyte identity; these waters are expected to be found in a solute excluding area [39, 40]. The waters that are being studied in the cases of *Eco*R1 and hemoglobin are either preferentially associated or dissociated with the protein or DNA of interest [35]. *Eco*RI is a model system for examining the role of hydration in protein-DNA interactions. *Eco*RI displays sequence-dependent hydration, thereby providing a means to investigate the role of hydration in protein-DNA interaction and catalysis [36]. *Eco*RI binds cognate DNA in the major groove through two α-helices via a series of hydrogen bonds, providing a direct readout between the protein and DNA [36]. However, when the specificity of *Eco*RI decreases, it develops the ability to cleave DNA at "star" sites [37]. In *Eco*RI, the rate of "star" sites increases when placed under increased osmotic pressure. These "star" sequences are similar to the recognition sites of *Eco*RI, suggesting that the increase in osmotic stress of the system decreases the level of specificity in binding, allowing *Eco*RI to bind and cleave these non-recognition sites [38, 40].

Another example of water playing a critical role in protein-DNA interactions is found in the *trp* repressor-operator complex in *E. coli*, where water mediates all polar interactions between the protein and DNA while still retaining sequence specificity [24, 37]. In situations when the concentration of tryptophan is too high, the *trp* repressor protein responds by repressing the transcription of specific genes [41]. Analysis of the crystal structure and osmotic stress binding experiments found about 75 water molecules in the protein-DNA interface of the *trp* repressor-operator complex. In the crystal structure, there are a number of water-mediated interactions between the protein and the phosphodiester backbone of DNA and several water-

mediated bonds between the protein and the DNA bases [41]. When bound to DNA, the *trp* repressor can make 1:1, 2:1, and 3:1 complexes. When examined under osmotic pressure, it was found that the 1:1 complex releases double the amount of water than those of the 2:1 and 3:1 complexes, which suggests that the 2:1 and 3:1 complexes retain more hydration than that of the 1:1 complex [41].

Another excellent example of a hydrated nonspecific protein-DNA complex in BamHI retains 120-150 waters [39]. Comparing the structures of BamHI to specific and nonspecific DNA shows drastic differences in the binding mechanisms. The BamHI-specific complex makes direct contacts between the DNA and protein. In contrast, the nonspecific complex contains a large space between the protein-DNA binding interface of the major groove, which can hold up to ~150 waters [40].

The RNA enzyme ncRNA, a hairpin ribozyme, is another example of a molecule that utilizes a highly ordered hydration network for structural dynamics and the formation of a catalytic core [42]. The catalytic core of the ncRNA is stabilized by a string of specific long-lived water molecules [42].

Preferential hydration is another aspect of the hydration of a protein's or DNA's surface, in which there is a layer of hydration that actively excludes other solutes or cosolvents from interacting with the surface of the protein or DNA [34, 36, 39]. Preferential hydration is the principle that interactions between the DNA and solutes, or the protein and solute are less energetically favorable than the interactions of DNA and water or protein and water [39].

1.2 Methods of analyzing hydration contributions to protein/DNA specificity

1.2.1 Osmotic stress

To understand the role of hydration in protein-DNA interactions, it is possible to alter water behavior by introducing neutral, inert osmolytes. These osmolytes increase the osmotic pressure while not interacting directly with the system [3, 33-35, 39]. Changes in hydration are then calculated by measuring the changes in binding affinity as a function of alterations to the behavior of bulk water- in other words, the chemical potential of water [38, 40]. By taking this approach, water is treated like a small molecule which determines how water interacts with specific structures and whether water is sequestered in cavities or in preferentially hydrated surfaces of the protein or DNA [34, 37]. This functions on the principle of there being less water in the system as the concentration of osmolyte increases, putting pressure on the system if the waters play an integral role. If the protein-DNA complex is thought to be a ‘dry’ complex, then increased osmotic stress will push the complex to the more dehydrated state and will show an increase in binding affinity [43]. However, if the complex requires the waters for proper structure formation, as the level of osmolyte grows, the amount of energy expended by the protein or DNA increases to hold on to those waters and is shown as a decrease in binding affinity [33, 36]. In other words, the addition of neutral osmolytes can be used to drive the system towards a less hydrated state [37, 38]. This change in energy can then be measured, which allows for the quantification of the net change of waters that are either taken up or released [3]. The use of osmotic pressure allows for the study of the water activity in solute inaccessible areas. Growing osmolyte concentration increases the pressure on the system and draws water out of normally inaccessible cavities, directing the complex to conform to a smaller volume [36]. Typically the formation of a protein-DNA complex is accompanied by a large decrease in hydration, which is

favored through osmotic stress [36]. Through osmotic pressure, it is possible to measure how changes in hydration affect binding, the equilibrium, and kinetics of the reaction to provide greater insight into the role of hydration in protein-DNA interactions [39].

Osmotic stress experiments provide a means of examining differences in hydration between different protein-DNA complexes without altering the system [40]. A linear change in binding affinity vs. osmotic pressure validates if the osmolytes act upon the hydration and not the complex itself. However, it is possible to increase the osmotic pressure of the system to the point that preferential hydration is perturbed, thus altering the system and acting upon the complex [40]. High levels of osmotic pressure should be avoided as a consequence, as it is imperative that the osmolytes act on the water alone; therefore, the difference between binding affinity and osmotic pressure must remain linear [40]. Osmolytes are naturally occurring and are frequently found in nature to keep proteins folded when conditions in the cell become less than favorable [44]. Osmolytes are used to examine the role of hydration and help stabilize the protein and prevent denaturation in some cases, such as dehydration [44]. Different osmolytes can have varying effects on the protein-DNA system- sucrose and glycerol tend to stabilize complexes, whereas urea can denature proteins at high concentrations [45]. These different properties of osmolytes are critical to know prior when determining what osmolytes will be used in hydration studies.

Another interesting aspect that can be examined by osmotic stress is in observing conformational changes in the protein or DNA. When significant structural changes occur, there is a corresponding change in the solvent-accessible surface area- complexes with less exposed surface area tend to be stabilized by neutral osmolytes [40, 41]. However, if waters play an important role in complex formation and remain bound to the protein, DNA, or protein-DNA

complex, increasing osmotic stress will lead to complex destabilization [41]. Cases where the identity of the osmolyte affects the system instead of showing consistent changes may be due to differences in the exposed surface area of the complex [39, 40]. By utilizing osmolytes of different sizes and charges, it is possible to determine if the observed changes are due to changes in hydration or are a function of osmolyte identity- if the results are in agreement independent of the osmolyte, it is likely the results are due to changes in hydration alone [39].

Previously, the restriction endonuclease *EcoR*I was examined under increasing osmotic pressure with both the high-affinity cognate sequence and the nonspecific sequence. Osmotic pressure was raised to the point where the nonspecific site could not compete against the high-affinity site [38]. These results were then used to establish the number of water molecules in the complex. The nonspecific complex retains ~110 more waters than its high-affinity variant. As the osmotic pressure increased, the high-affinity complex was stabilized due to the lack of water-mediated interactions. In contrast, the nonspecific complex was destabilized due to the energetic cost of holding onto the waters under increasing strain [38-40]. In the *Eco*RI study, various osmolytes were utilized with the same trend, proving that the change in hydration is due to the increased strain on retaining the waters rather than the identity of the osmolyte used [38].

Water has been shown to play essential roles in biological processes necessary for life, including protein-DNA interactions and folding of cellular structures like the cellular membrane, proteins, and DNA [33]. How the system excludes solutes is of great interest for understanding the system as a whole- solutes can be excluded via steric exclusion or through preferential hydration. In steric exclusion, there are cavities in proteins and DNA that are small enough to keep solutes out but large enough to admit water [33]. Osmolytes of different sizes are extremely useful in studying water locations, as they will interact with different regions in the system. A

small osmolyte may enter a cavity that a larger osmolyte has no access to, providing valuable information on the location of water [33]. Preferential hydration takes into account the surface of the protein or DNA and looks at the ratio of solute to water. There will be more water molecules in areas of preferential hydration than solute as the surface makes direct interactions with the water, which excludes solutes. To utilize osmotic stress, it is imperative that the osmolyte not interact directly with the protein or DNA so that any observed effects come from the change in pressure due to the osmolyte concentration. As varying osmolytes will affect the system differently in terms of viscosity, water pressure, osmotic pressure, and dielectric, it is essential to know that the changes in the system are not due to these properties [37]. A method to test that the results are not from the identity of the osmolyte is to test the system utilizing a variety of osmolytes of different sizes and chemical compositions. If the results are consistent across different osmolytes, then the results are due to changes in osmotic pressure and not due to the osmolyte. This can also be displayed when plotting the results- any changes due to osmotic stress will be linear on the logarithmic scale; if the line is nonlinear, then it suggests that the osmolyte is directly interacting with the system or the concentration of the osmolyte has increased to a level that affects the preferential waters on the surface of the system [33, 35]. If there is a change between different osmolytes, this could be characteristic of a change in available surface area [35]. An issue associated with the use of osmotic stress experiments is that the results only provide the net water uptake or release but provide no information about the location of the change in hydration. Osmotic stress is beneficial in exploring the changes in the water activity and provides an overall picture of how hydration affects protein-DNA specificity.

Osmotic stress experiments provide insight into the hydration of protein-DNA complexes by perturbing the complex as osmotic pressure increases. However, osmotic stress can only tell a

change in hydration occurs and the magnitude of the change. It does not provide detail about the location of the changes in hydration and whether the water is found in the protein, DNA, or both [41]. The study of osmotic stress is interesting as this process examines the role of waters included rather than excluded when the solutes are unable to interact with the system [35]. Studying a system under increased osmotic pressure helps to understand the role of hydration in sequence selectivity, aiding in understanding how differential hydration affects specificity [32].

1.2.2 Volumetrics

There are many techniques used to study hydration- X-ray crystallography provides detail on the location of ordered waters, whereas NMR gives information on the dynamics of water [46, 47]. However, crystallography may miss the dynamic waters, and NMR likewise has difficulty detecting rapidly exchanging waters [46, 47]. By utilizing a high-precision densimeter to analyze the partial molar volume of the sample, it is possible to obtain volumetric data that pertains to the whole hydration shell of the system. This, in turn, provides a thermodynamic basis to understand the role of hydration in biomolecular interactions [47]. It is important to remember that although partial molar volumes offer an excellent idea to the role of hydration, it still does not provide any insight into the location or dynamics of the hydration shell. Previously, Chalikian determined that the apparent molar volume of duplex DNA is found between 152.0 to 186.6 cm^3mol^{-1}, which provides a reference when examining other duplex DNA [46]. This does not explain how the apparent molar volume of DNA may change when bound to protein.

Proteins and DNA have an associated volume from the tertiary structure comprised of the space taken up by atoms, cavities, and the hydration shell [48, 49]. Waters in the shell of hydration are found around the protein, DNA, or complex and are chemically distinct from bulk solution, being denser, more ordered, and less mobile [46, 47]. The hydration shell is affected by

electrostatic interactions and the amount of hydrophilic polar groups on the surface [46]. The specific volume of protein and DNA can change due to interactions with other molecules, such as when protein interacts with DNA to result in either the uptake or the release of water from/to bulk solution, which is a key aspect of volume that is important to examine [47, 49]. The level of hydration of protein and DNA is also affected by changes in conformation, which may be due to folding, denaturing, or structural shifts due to binding interactions [50]. Folded proteins tend to have a larger specific volume than unfolded or denatured proteins [50]. DNA duplexes are highly hydrated and contain differing hydration patterns depending on structure and conformation; single-stranded DNA is more exposed and may be slightly more hydrated than the duplex counterpart. However, hydration in duplex DNA is more restricted, suggesting a structural role played by the hydration shell [51]. Understanding the patterns and effects of hydration on the properties of proteins and DNA will allow for a greater understanding of how these patterns are altered due to binding events [52]. This understanding of the role of hydration makes it possible to learn more about the thermodynamics of how protein-DNA complexes achieve specificity [47, 53]. Specific hydration patterns have been proven to have a significant effect on protein-DNA specificity and affinity, as displayed by the *trp* repressor complex [53, 54].

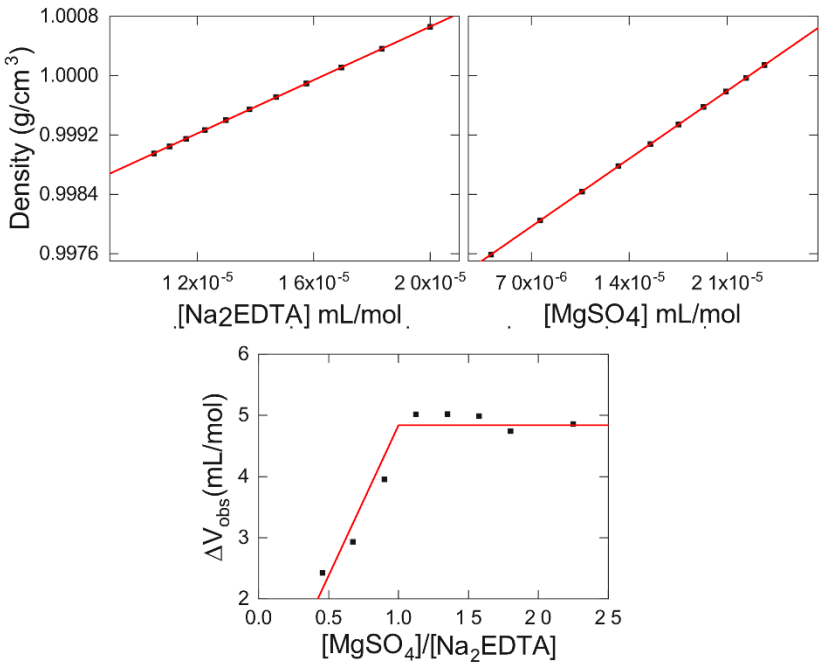


Figure 1-2 Titration of Na_2EDTA with $MgSO_4$

To determine the change in volume of a system, the change in volume of the individual components must first be determined by measuring the solution's density. Both Na_2EDTA and $MgSO_4$ were dissolved in ultrapure H_2O and were examined by densimetry to determine the partial specific volume individually. Once the individual elements have been examined, the titration is performed where the amount of Na_2EDTA remains constant. An increasing concentration of $MgSO_4$ is added; as the $MgSO_4$ is titrated into the sample, a complex is formed. For a successful titration to be completed, it is essential to show that the system has reached saturation; in this case, saturation was determined to be before a molar ratio of 2. Saturation was successful, as seen by the single transition point at a 1 molar ratio and the consistent change in volume that follows. The titration is straightforward to determine the change in volume upon forming the 1:1 complex, as this is the one inflection point in the titration.

Measuring volume changes makes it possible to learn more about protein-DNA recognition through changes in hydration of combined elements versus single elements [49]. A densimeter can be used to calculate the partial molar volume of a sample and is accurate up to six decimal places. By directly measuring the density of the sample, it is then possible to convert density into the change in observed volume, as shown in equation 1-4.

$$\Delta V_{obs} = \frac{(\rho - \rho_0) - (\rho' - \rho_0')(1 + \frac{V'}{V_0'})}{\rho_0 c} \qquad \textit{Equation 1-4}$$

Where ΔV_{obs} is the observed change in volume of the complex, p is the density of the DNA in buffer, p_o is the density of the buffer, p' is the density of the complex in this case protein bound to DNA, p_o' is the density of the protein in buffer, V' is the total volume of protein added to the sample, V_o' is the initial volume, and c is the initial DNA concentration, a dilution correction is included in this equation [55]. An example of calculating the change in observed volume of a complex can be seen in Figure 1-2, where $MgSO_4$ was titrated into EDTA in molar ratios from zero to just over two. Partial specific molar volume is another method using a densimeter to determine the levels of hydration in a complex (Figure 1-3). Where the partial specific molar volume (V_o) of each sample is calculated individually utilizing the slope of the dilution of the sample (density vs. concentration), the molar mass of the sample (m), and the density of the buffer (P_o) as seen in equation 1-5 [46, 56]:

$$V_o = \frac{slope - m}{P_o}$$ Equation 1-5

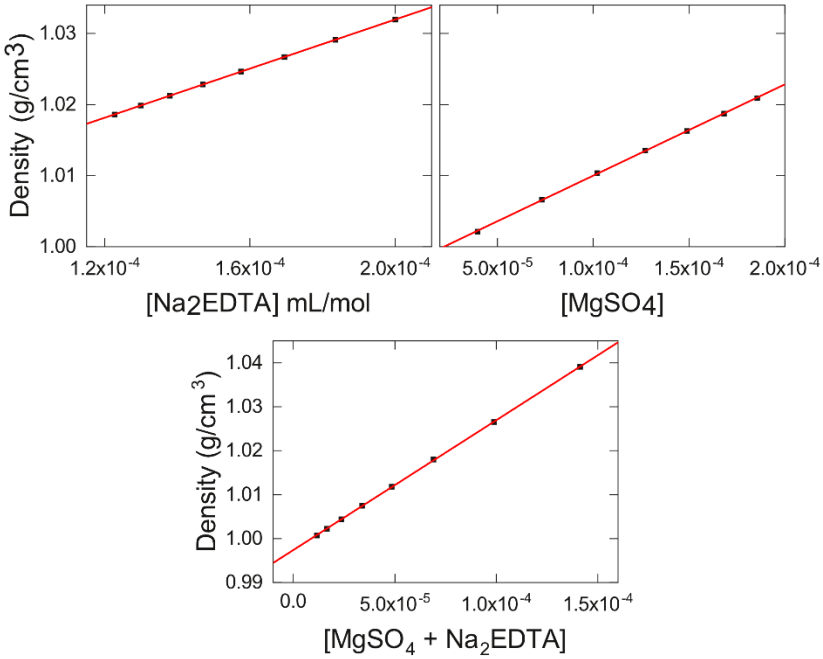


Figure 1-3 Dilution of a 1:1 molar ratio of Na_2EDTA with $MgSO_4$

Another method of determining the partial specific molar volume of a complex is by performing a dilution experiment. In which the complex of interest is formed from the start and is serially diluted through the addition of buffer. A dilution of the complex's individual components is performed to check that the system is working correctly.

The changes in volume and protein hydration differ depending on the interaction type, such as between high and low-affinity interactions (reaction volume). Measuring the difference in reaction density using a vibrating tube densitometer provides a means to calculate the change in partial molar volume (ΔV_{obs}). Observing the change in volume is helpful since the volume affects the whole molecule or complex, like when examining the waters of hydration rather than populations that will only affect specific areas or are present under particular conditions [53].

1.2.3 Isothermal Titration Calorimetry

Identifying how protein and DNA interact is essential for understanding biological function. The folding and unfolding of proteins, the association, and disassociation of protein-DNA complexes result in a change in enthalpy (ΔH) due to the changes in thermodynamics from the forming and breaking of bonds [57]. The change in enthalpy is directly measurable through isothermal titration calorimetry (ITC), which provides a complete thermodynamic profile by measuring the change in enthalpy (ΔH), binding affinity (K_a), and the binding stoichiometry (n) through the titration of one substance into another [58, 59]. By directly calculating the change in enthalpy (ΔH), and binding affinity (K_a) of the complex, it is possible to calculate the change in Gibbs free energy (ΔG) and the change in entropy $(T\Delta S)$ of the system (equation 1-6), which indicates what conditions directly affect protein-DNA specificity [59].

$$\Delta G^\circ = -RT \ln K_{eq} = \Delta H - T\Delta S^\circ \qquad \text{Equation 1-6}$$

A system that utilized protein-protein interaction to better understand specificity employed the signaling pathways involving the Src homology 2 (SH2) domain and proteins with a phosphorylated tyrosine. SH2 domains are highly prevalent in cytoplasmic proteins and are highly conserved in both sequence and structure [57]. Which leads to the question of what drives the specificity of SH2 domains? By exploring the thermodynamic properties of specific and nonspecific complexes with tyrosyl phosphopeptides, it is possible to elucidate what is responsible for specificity. Analysis of structural data revealed that the specific complex involves multiple amino acids interacting with the peptide and the SH2 domain. In contrast, the nonspecific complex is more dynamic and posses more degrees of freedom, displayed by the increased entropy in the nonspecific complex. The determination of Gibbs free energy through

ITC has shown a difference of greater than 8 kJ mol^{-1} between specific and non-specific interactions, suggesting that nonspecific complexes have lower levels of Gibbs free energy but a more significant increase in the entropy of the system [57].

Another method for utilizing ITC to explore specificity in protein-DNA interactions is by calculating the change in heat capacity (ΔCp). To form specific protein-DNA complexes, the protein binds to DNA by recognizing a specific DNA sequence, while nonspecific complexes will bind to DNA without recognizing the sequence. The specific complex usually binds with high-affinity interactions, and the nonspecific complex tends to have weaker interactions. Calculating the (ΔCp) is done by plotting the change in enthalpy vs the temperature change; specific complexes are expected to have a large and negative (ΔCp). In contrast, nonspecific complexes will have a small to no (ΔCp) [57]. When plotting (ΔCp), the final plot is expected to have a linear trend; a nonlinear trend can result if the protein is binding both specifically and nonspecifically [60].

$$(\Delta C_P = (\frac{\partial \Delta H}{\partial T})_P = T(\frac{\partial \Delta S}{\partial T})_P) \qquad \text{Equation 1-7}$$

The differences in (ΔCp) may be due to changes in the solvent-exposed area, where the specific interaction displaces the solvent upon binding [57, 61]. One method of determining the thermodynamic properties between equilibrium states is by comparing structural information to the (ΔCp) - with detailed structural information. It is possible to determine whether the (ΔCp) is due to changes in the solvent-exposed surface area, demonstrating the effect of hydration on (ΔCp) [61]. Previously it was shown that hydrophobic or non-polar surface areas play a strong role in the (ΔCp), whereas hydrophilic or polar surface areas reduce (ΔCp) [61]. A prominent example of protein-DNA interactions studied by ITC in conjunction with the crystal structure is

the *trp* repressor protein bound to a 20 base-pair DNA containing the cognate binding site [61]. The *trp* repressor complex is well hydrated, with multiple water-mediated contacts between the protein and DNA required for the specific binding. The water molecules in the specific *trp* repressor complex contribute to the (ΔCp), making it more negative than expected compared with the (ΔCp) calculated from the burial of hydrophobic surface area [61]. Previously, it was believed that the addition of water in the protein-DNA interface would have a negative impact on the (ΔCp) due to the entropic cost of having to restrain the water in a specific location. However, the water also reduces the potential of having noncovalent interactions, thereby increasing the enthalpic gain to offset the entropic cost [61].

Every system has a unique thermodynamic profile due to how solvents interact with biomolecules and how biomolecules interact with each other. But there are trends that can be explored- using the thermodynamic profiles of protein-DNA interactions, it is possible to understand better what is required for specificity. In this context, the focus is on protein-DNA interactions; ITC is a highly versatile technique and can be utilized to study various interactions, including protein-protein, protein-peptide, hormone-receptor, antibody-antigen, and enzyme-substrate interactions [57]. However, ITC is not limited to only interactions between molecules- ITC can also be used to study the effects of pH changes by titrating in buffers of increasing or decreasing pH [57].

To measure the enthalpy change in the system, the ITC utilizes two cells- the reference cell contains only buffer, and the sample cell contains one of the biomolecules being examined, such as DNA. The second biomolecule, protein, is then titrated into the sample cell in equal aliquots and the resulting interaction is either exothermic or endothermic. The temperature of the reference cell is then adjusted to match the sample cell. For exothermic binding, the sample cell

will lose heat; if endothermic, the sample cell will gain heat [57, 59]. This experiment aims to saturate the DNA binding sites, so any change in heat at the end of the titration is due to binding rather than dilution [57, 59]. A complete ITC titration provides enough information to determine the change in enthalpy (ΔH), binding affinity (K_a), and stoichiometry (n) [57, 59].

Determining the stoichiometry of the system is a simple task once the titration is completed. The final curve in which the change in heat is plotted against the molar ratio provides a clear visualization of the stoichiometric preferences. In the event of a single transition site, the binding is 1:1, while with more than one transition, then a 2:1 complex is likely with, distinct (ΔH) and (K_a) for each transition [62].

1.3 Transcription factors

1.3.1 ETS family of transcription factors

To study the role of hydration in protein-DNA specificity, we have chosen the ETS (E26 transformation-specific) family of transcription factors, a transcription factor family found throughout the animal kingdom, including the lower metazoans [58, 63-65]. The conserved genes that signal the development of the ETS family proteins are proto-oncogenes and are essential for the proper development of living creatures as they signal for specific events to take place in cellular development [63, 66]. The ETS family proto-oncogenes are one of the most conserved families and can be found in sea urchins, *Drosophila*, and more complex organisms up to and including humans [63, 67]. It was previously found that different ETS transcription factors regulate housekeeping genes in various tissues, which are found in all cell types and are essential for the health of the cell [67, 68]. The family acts both as activators and repressors of transcription, and in humans has been found to play key roles in angiogenesis, hematopoiesis, myogenesis, neuronal development, and ossification [69-71].

The ETS family consists of 28 members in the human genome. All of which contain a structurally conserved winged helix-turn-helix DNA binding domain (DBD) that is primary sequence divergent (85-amino acid residues in length) and binds to a purine-rich cognate sequence containing 5'-GGA-3' [32, 58, 63, 64, 66, 67, 69, 70, 72-77]. Members of the family can be highly divergent, having as low as ~30% sequence similarity and as high as 97% to the founding member Ets-1. ETS family members recognize a conserved DNA motif with similar binding despite drastic changes in the primary amino acid sequence [69, 76]. The few residues that are highly conserved throughout the ETS family tend to be residues that make direct interactions between the protein and DNA [72]. The ETS family of transcription factors act by binding DNA at target sequences to drive the expression of specific genes [67]. ETS family members bind DNA sequences specifically up to ~9 base pairs long but make nonspecific contacts to an additional 3-6 bases [67]. ETS members interact with DNA through direct and indirect interactions based on the DNA sequence and the shape of the phosphodiester backbone [67]. Specific interactions are achieved by binding the recognition helix (H3) into the major groove of DNA- within H3 are two conserved arginines that contact the 5'-GG-3' of the consensus sequence [64, 67, 69, 72]. Upon binding DNA, ETS proteins bend the DNA between 11-28°, and the residues flanking the recognition helix stabilize the complex by interacting with the minor groove phosphate backbone [64, 69]. Previous studies have found that some members like PU.1 make few direct contacts to DNA outside of the consensus sequence and rely heavily on water-mediated hydrogen bonds, hydrophobic and electrostatic interactions. Other ETS members make only a few direct interactions with the flanking regions of DNA, usually in the wing region to the phosphate backbone in the minor groove [67, 69]. Alterations in the flanking region of DNA can result in over a 400 fold change in binding affinity [67]. The sequence

selectivity, DNA flexibility, and the types of interactions displayed in the crystal structure complexes of ETS family members suggest that DNA recognition is achieved through a combination of direct and indirect readout of the DNA [69].

Within the ETS family, there are four classes: class I is the largest and binds the consensus sequence 5'-ACCGGAAGT-3', class II binds a consensus region containing the sequence 5'-CCCGGAAGT-3', class III consists of the SPI subfamily and binds the consensus region that is adenine-rich containing the 5'-GGA-3', and finally, class IV prefers 5'-GGAT-3' rather than 5'-GGAA-3' [64, 67, 70, 78]. Within the ETS family, there is binding redundancy with different sequences, especially in the case of housekeeping genes [68, 74]. Previous experiments with Ets-1 by chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) experiments show that both Ets-1 and GABPA bind the same promotors of housekeeping genes, but Ets-1 retains specificity and sole occupancy of the gene enhancers for T-cells [68]. Previous research has suggested that ETS family members determine the binding site based on DNA flanking regions rather than the core consensus sequence [69].

Although ETS family members are structurally conserved and bind cognate DNA sequences with similar binding affinities, they are not functionally interchangeable [64, 66, 70, 72, 79]. Within the ETS family of transcription factors, multiple subfamilies retain the proper folding and bind the cognate sequence with similar binding affinities but have distinctive biochemical responses [67]. Ets-1, for example, is involved in the activation of T cells and specifically binds enhancer and promotor sites for T-cell activation [68]. Different ETS family members have different domains outside of the ETS domain, including the PNT and OST domains [80, 81]. The PNT domain is a structural domain found on multiple ETS family proteins, whereas the OST domain assists in cofactor recruitment for the ETS family member

GABPA [67, 80, 81]. Between any structured domains in ETS family members are intrinsically disordered regions, which can interact with other proteins or DNA and can be the site of post-translational modifications [67]. Several ETS family members have structured C-terminal helices. In the case of GABPA, the C-terminal helices assist in binding GABPβ to improve binding affinity, while the C-terminal helices act as an autoinhibitory domain in ETV6 [67]. Ets-1 has two N-terminal helices (HI1 and HI2) that interact with C-terminal helices H4 and H5 to allosterically autoinhibit DNA binding [67]. These autoinhibitory domains found in multiple ETS family members are essential for proper function- the cancerous mutation of Ets-1 in the E26 retrovirus has no C-terminal inhibition domain, resulting in increased DNA-binding that is believed to increase the probability of cancerous cells [67]. Phosphorylation also acts as an additional autoinhibition mechanism for Ets-1, while interaction with certain binding partners can alleviate autoinhibition.

These differences between ETS family members allow each member to serve different biological functions within the cell [67]. However, many ETS family members are found co-expressed within the same cell, or tissue type, with up to half of the ETS family members found in the majority of tissue types examined [67, 74]. Members of the ETS family act as transcriptional activators and many members of the ETS family have been found to cause various cancers when overexpressed [69, 76]. Specific tissue types may also highly express individual ETS family members, such as SPDEF and ESE in mammary tumors and ERG with PEA3 in prostate tumors [67, 78]. Since ETS family members bind very similar DNA sequences, small mutations and translocations in the DNA binding sites can lead to the expression of the wrong family member, possibly leading to the development of various cancers [69, 70]. Translocations can lead to the loss of the DNA binding domain in ETS family members or lead to the protein's

activation and over-expression [70]. Understanding protein-DNA specificity and the regulation of transcription is essential for understanding the development of these cancers and for potential therapeutic treatments [69].

The ETS family member PU.1 binds specific DNA sites to cause cells to differentiate into hematopoietic stem cells, monocytic and B lymphocytic cells, and T cells [64, 67, 82, 83]. Even though family members serve different biological functions, they are often co-expressed in various tissue types and are often linked with a variety of cancers. In Ets-1 and PU.1, they tend to drive differentiation in opposite directions- T cell differentiation results when PU.1 activity decreases and Ets-1 activity increases [64, 66, 67]. That many ETS family members are co-expressed in the same cells and bind to DNA containing identical cognate sequences with similar affinities leads to the question of how the cell differentiates between the proteins and how the proteins determine specificity in the binding sites.

1.3.2 PU.1

ETS family transcription factor PU.1 can bind specifically to more than 70 different sequences containing a core consensus sequence (5'-GGAA-3') with differing flanking regions. Variations in the flanking regions are responsible for the spread of binding affinities that can be as large as over a 400-fold difference [20, 73, 84, 85]. Of the ETS family members, PU.1 is the most sequence divergent from its predecessors and is a master regulator of genes in myeloid cells [85, 86]. When compared with one of the earliest family members, Ets-1, PU.1 is the most sequence divergent, sharing only ~30% sequence homology (Figure 1-4). Despite this variation,

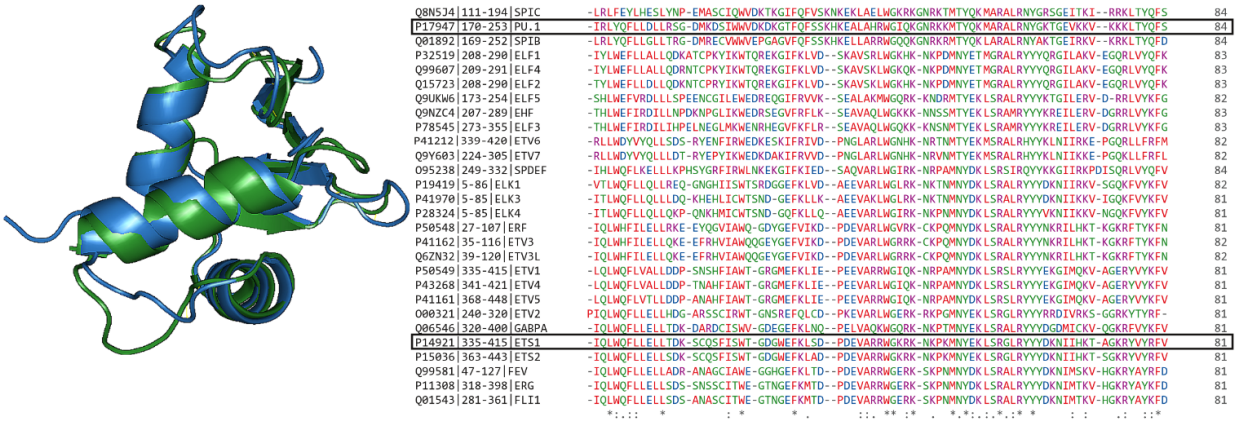


Figure 1-4 Comparison of the tertiary structure of PU.1 vs. Ets-1 and the primary sequence of the DNA binding domain of all ETS family members.

A) *A structural alignment of the ETS domains of PU.1 and Ets-1 shows the conserved structure of the DNA binding domain. The structured elements are highly conserved between the family members, with the helixes and beta sheets overlaying nearly perfectly. B) Through a sequence alignment of the ETS domain of the entire ETS family of transcription factors, it is possible to determine the level of conservation of the primary sequence. By comparing the most evolutionarily distant member, Ets-1, it was determined that PU.1 is the most divergent in the primary sequence. PU.1 and Ets-1 are the most evolutionarily distant members of the ETS family. The retention of the common structure of the DBD paired with the different primary sequence provides a method to design a series of chimeric constructs in which specific PU.1 secondary structures are exchanged with the corresponding structure of Ets-1 individually, as a means of determining the role of hydration in the PU.1 high-affinity complex.*

the structurally conserved DNA binding (ETS) domain in both PU.1 and Ets-1 is found in the C-terminus of the protein (Figure 1-3) [66, 73, 79]. The site selectivity of PU.1 for DNA is of

utmost importance as PU.1 alone or in complex with other transcription factors regulates at a minimum 24 genes essential for proper immunological and developmental functions. PU.1 is a master regulator of hematopoietic cell lines and regulates macrophages, monocytes, and lymphocytes in B-cell development [20, 65, 66, 72, 77, 87, 88]. PU.1 is a pioneer transcription factor, as unlike Ets-1 and other ETS family members, PU.1 is able to bind to a diverse range of target sites, has more robust sequence selectivity, and is found at a higher rate in osmotically sensitive genes [64, 66, 72, 73, 79]. PU.1 is responsible for directing cell differentiation of hematopoietic stem cells and determines cell fate and functions [64, 65]. Within cells, PU.1 acts to create enhancers for gene expression and can form complexes with IRF8 to achieve stronger binding affinities [86]. When compared with Ets-1, PU.1 acts in reverse- in T-cell differentiation, Ets-1 increases activity while PU.1 decreases [64]. When improperly expressed, PU.1 has been found to lead to several diseases and cancers, including rheumatism, Alzheimer's disease, and hematologic cancers [64, 65, 88]. PU.1 is also implicated in the development of acute myeloid leukemia (AML), and has been found to cause AML when levels of PU.1 are decreased in the cell [83].

PU.1 contains the ability to bind with methylated DNA, can bind to chromatin that is unreachable to DNase I, and has the capacity for cooperative recruitment of other transcription factors [66, 73, 79, 87]. When PU.1 and Ets-1 were examined on hemimethylated DNA, they had very different responses. PU.1 with hemimethylated DNA on the forward sequence showed a decrease in binding affinity by 5-fold, whereas the hemimethylated DNA on the reverse sequence decreased binding affinity over 200-fold [64, 79]. Ets-1 could not noticeably bind the forward methylated sequence but only had a slight drop in binding affinity upon binding the methylated reverse sequence [79]. When examining fully methylated DNA sequences, PU.1

bound with only a 12-fold decrease in binding affinity, with Ets-1 having a reduction by 160-fold [64, 79]. The level of hydration of PU.1 allows for compensation when bound to methylated DNA, leading to stronger binding affinities than Ets-1 [64, 79]. In ETS family members, there is a highly conserved arginine responsible for contacting the DNA site. In PU.1, this arginine makes three water-mediated interactions, but in Ets-1, the same arginine only makes one water-mediated interaction [79]. This conserved arginine is found next to the methylated cytosine on the forward sequence. Due to there being less hydration in the Ets-1/DNA interface, is likely unable to compensate for the less-favorable interactions of the methylated base [79]. PU.1 can also recruit a variety of cofactors and transcription factors depending on the specific cell type [87].

Thermodynamically, Ets-1 and PU.1 behave in very different manners. Comparing heat capacity (C_p) of Ets-1 and PU.1 shows Ets-1 has a larger ΔC_p and PU.1 has a smaller ΔC_p; this is expected due to the hydration of the PU.1/DNA interface [66]. On the flip side, PU.1 has increased stringency of site selectivity and forms complexes with DNA at ~100 times slower than Ets-1, but the PU.1/DNA complex is longer-lived than the Ets-1/DNA complex [64, 66]. Given the body of evidence, it is believed PU.1 binding requires a proper hydration network rather than direct interactions between the protein and DNA [66]. PU.1 was also found to be more dynamic in site discrimination than Ets-1 when examined by dynamic light scattering. When the level of dynamics was compared with the data from DNA footprinting, it was understood that the dynamic parts of the PU.1/DNA complex are found outside of the recognition helix and major groove of DNA [73].

PU.1 interacts with a 10-basepair stretch making up the recognition site, and upon binding, DNA bends by 8° [77, 78]. Within the DNA binding domain, there are four highly

conserved amino acid residues- two arginines that are located in the recognition helix that interacts with the consensus sequence, a lysine located in the wing that binds to the 5' flanking region in the minor groove, and a second lysine forms a salt bridge on the opposite side in the turn and contacts the 3' flanking region of DNA of the consensus site [77]. To form the protein/DNA complex, the recognition helix (H3) of PU.1 is inserted into the major groove and makes a series of direct and water-mediated interactions [72, 84, 85].

PU.1 obtains sequence selectivity through both direct and indirect readout. The recognition helix (H3) utilizes water to make indirect contacts with specific bases of DNA. The 8° curvature upon binding allows PU.1 to make contacts with the phosphate backbone of the flanking regions of the DNA. This increases the prevalence of indirect readout due to the absence of base-specific readout [20, 64, 84, 85]. The bend in DNA does not correspond to a significant structural change in PU.1 itself, as examined by NMR and CD experiments [84]. The binding affinity of PU.1 to DNA is highly dependent on the identity of the flanking regions around the consensus sequence, which can dynamically reduce the flexibility of PU.1 to stabilize the protein/DNA complex [84]. PU.1 binding affinity is also more sensitive to changes in the flanking regions upstream (5' side) of the consensus sequence [84]. When examined under increasing temperatures, it was discovered that sequences with higher affinities were more strongly affected by changes in temperature, displaying lower affinities and more negative ΔC_p [84]. Likewise, sequences that were already weaker binding sites were less affected by the changing temperature and had smaller changes in ΔC_p [84]. The thermodynamic profile of PU.1 binding to DNAs with differing flanking regions shows that less favorable interactions are entropically stabilized, whereas the more favorable bindings are enthalpically stabilized [84]. PU.1 also displays a strong salt dependence when bound to differing sequences. The highest

affinity sites have a more negligible salt dependence, and lower affinity sites have a stronger salt dependence [84].

When examining the crystal structure of PU.1 bound to DNA (1PUE), 27 ordered solvent molecules around the DNA were found, with water in the major groove bridging the protein and DNA [64, 77]. The highly conserved arginines in the recognition helix interact with DNA via direct and water-mediated interactions [77]. Three non-conserved residues form well-ordered water-mediated contacts between the DNA and protein, including Thr 226, Gln 228, and Asn 236 [77]. The co-crystal structure of Ets-1 bound to high-affinity DNA is barely hydrated, especially in comparison with the PU.1 co-crystal complex [64].

Structurally, the PU.1 high and low-affinity complexes are homologous when studied by DNase I and hydroxyl radical footprinting. However, there is a divergence in the level of protection obtained by the differing complexes [73]. Dimethyl sulfate footprinting has proven that the core consensus is protected differently between the complexes. The high-affinity complex is not accessible to dimethyl sulfate, and the low-affinity complex is accessible [73, 85]. The significant difference between the high-affinity and low-affinity complexes is the level of hydration. The high-affinity complex utilizes multiple water-mediated contacts between H3 and the core consensus sequence, and these contacts are not present in the low-affinity complex. This suggests water acts to protect the binding site from dimethyl sulfate and helps increase site selectivity [85]. In terms of salt dependence, the high-affinity PU.1 complex is destabilized by increasing salt concentrations but not as drastically affected compared to the low-affinity complex [66, 73]. Ets-1 has no distinction in salt dependence relative to sequence, and the dependence scales appropriately with the number of phosphate contacts within the sequence [73]. The trend with salt dependence is inverse to the dependence with hydration, where the

PU.1/high-affinity complex is greatly destabilized by increased osmotic pressure, and the low-affinity complex is stabilized [66].

Previous work by both CD and NMR has shown that the selectivity of PU.1 is not from coupled folding. The secondary and tertiary structures of PU.1 in the bound and unbound states are nearly identical. The calculated ΔC_p was small, suggesting that coupled folding is not a driving factor for PU.1 sequence selectivity [32].

DNA binding experiments between PU.1 with high-affinity DNA vs. Ets-1 with high-affinity DNA have shown different binding stoichiometries. PU.1 makes a 2:1 protein/DNA complex at high concentrations, while Ets-1 on high-affinity DNA remains a 1:1 complex [66]. PU.1 binds the 10 base pair sequence in a monomeric fashion but requires longer DNA sequences to form the 2:1 complex [65]. Through NMR and hydroxyl radical DNA footprinting, the 2:1 complex of PU.1 on DNA has been confirmed. The second PU.1 is thought to bind to the distal side of the first PU.1 protein, away from the protein-DNA interface with a potential conformational change [65]. Dimeric PU.1 in bound and unbound states were found to have distinct conformations [88]. This dimeric binding of PU.1 to DNA is thought to be a possible mechanism for autoinhibition of the PU.1 ETS domain on DNA. As full-length Ets-1 has an autoinhibitory domain, PU.1 may have adopted a dimeric form to act as an autoinhibitory mechanism due to the lack of an actual autoinhibitory domain [65, 88].

1.3.3 Hydration of PU.1

The recognition of DNA by PU.1 is sensitive to changes in the environment, specifically in changes in osmotic stress [72]. Examining the crystal structure of PU.1 bound to DNA, several waters were found suggesting a structural application for their presence [64, 66, 85]. In the PU.1 co-crystal structure with DNA, most contacts are water-mediated contacts, whereas, in Ets-1, the

interface is largely dehydrated [64, 66, 72]. When tested under increased osmotic pressure, the binding affinity of PU.1 to high-affinity DNA drastically decreases as osmotic pressure increases, whereas the Ets-1/high-affinity complex displays no sensitivity to osmotic pressure [64, 66, 73, 85]. Previous studies show that when bound to high-affinity DNA, there is a net uptake of water upon complex formation, while the PU.1/low-affinity complex and the Ets-1/high-affinity complex are hydration neutral [64, 66, 73, 85]. The PU.1 high and low-affinity complexes retain the same binding structure, as examined by DNase I foot printing and hydroxyl radical foot printing, despite the drastically different hydration levels [73]. When tested by dimethyl sulfate footprinting, a solute-excluding cavity was discovered in the PU.1/high-affinity complex that is not present in the low-affinity complex [85]. It is hypothesized that PU.1 discriminates sequences based on hydration, needing specific waters to bind and create base-specific contacts along the backbone of the flanking regions. More base-specific interactions in the inaccessible cavity were found around the 5'-GGAA'3- sequence in the high-affinity complex [85]. This water-accessible cavity is located only in the high-affinity complex, which explains why the low-affinity complex is unaffected by the increase in osmotic pressure [85]. DMS footprinting clearly shows that the guanosines in the major groove 5'-GGAA-3' of the high-affinity complex are protected, but these same guanosines in the low-affinity complex are not protected [85]. The PU.1 high-affinity complex was initially expected to release waters upon formation of the complex since to achieve high-affinity interactions, there is generally a sizeable complementary surface that excludes waters, but previous research has shown that the waters in the PU.1/DNA complex are essential and are retained upon forming the complex [85]. The use of osmotic pressure is ideal for studying hydration as the addition of the neutral osmolytes drives

the formation of the protein/DNA complex due to there being less water, thus making the complex more dehydrated [85].

To understand how binding affinity (K_B) is affected by a change in osmotic pressure, we studied how the system responds to changes in osmolality by adjusting the concentration of the neutral osmolyte glycine betaine. We can utilize this method as osmotic pressure (π) and water activity (a_w) correspond to osmolality (Osm) where [66, 85]:

$$Osm = \frac{\pi}{RT} = -55.5 \ln a_w \qquad \textit{Equation 1-8}$$

Using equation 1-9, it is possible to calculate the change in preferential hydration ($\Delta\Gamma_{PW}$) between unbound PU.1 and the complex. If osmolality is negative, then there is a release of water due to the complex being stabilized by the osmolyte, a positive osmolality shows a net uptake of water due to a destabilization of the complex, and if there is no change, then hydration between the bound and unbound complexes is net neutral [66, 85].

$$\frac{\partial \ln K_B}{\partial Osm} = \frac{\partial \ln K_B}{-55.5 \partial \ln a_w} = -\frac{\Lambda T_{pw}}{55.5} \qquad \textit{Equation 1-9}$$

Using equation 1-9, PU.1, with a variety of neutral osmolytes, was found to uptake water in a high-affinity complex, was hydration neutral in a low-affinity complex, and released water when bound non-specifically [85]. PU.1/high-affinity DNA interactions are of great interest as the formation of high-affinity complexes are usually accompanied by a release of water; the inclusion of osmolyte into the binding mixture minimizes the amount of surface exposed to solvent, making the release of water energetically favorable [85]. PU.1 holds onto the waters

when binding high-affinity sequences. When osmotic pressure is increased enough that the water contacts are disrupted, the binding affinity decreases. This shows that to form a high-affinity complex, PU.1 must uptake waters to mediate the contacts [85].

Tyrosine 252 in PU.1, a highly conserved residue found in 26 of the 28 ETS family members, forms a water-mediated contact between itself, arginine 235, and the backbone phosphate of a DNA base after the 5'-GGAA-3' [72, 85]. The water-mediated contact with tyrosine is the last of a string of hydrated interactions in PU.1. This includes several water-mediated contacts throughout the protein/DNA interface, such as Thr 226 and Glu 228, and in the flanking regions Asn 221, Lys 223, and Lys 229 [85]. Previously our lab has examined the role of hydration in the PU.1 high-affinity complex by mutating the tyrosine to phenylalanine, thereby removing the ability to form the water-mediated interaction [72]. Tyrosine to phenylalanine was done in both PU.1 and Ets-1 to understand how this interaction affects binding affinity and hydration in the respective high-affinity complexes [72]. Under osmotic stress, the -OH loss caused a decrease in PU.1's sensitivity to increased osmotic pressure by ~25% [72]. Ets-1 with the mutation saw a ~60 fold decrease in binding affinity under normo-osmotic conditions and had a 35% more positive slope showing that the complex was more dehydrated than with wildtype (WT) Ets-1 [72]. When examined by CD, both mutants showed decreases in β-sheet content, but the melting temperature of the Ets-1 mutant was ~8°C higher than WT Ets-1 [72]. Mutant PU.1 was denatured and refolded and had a melting point similar to WT PU.1 [72]. The change in binding affinity of Ets-1 with the mutation appeared to be due to a change in local dynamics rather than due to a change in hydration [72]. This study was the first to prove that by mutating critical residues in the protein, it is possible to alter hydration in specific complexes and led to the idea that hydration in PU.1 specifically evolved as a means of

sequence selectivity [72]. This led directly to us asking whether it is possible to completely knock out the hydration dependence in the PU.1/high-affinity complex and if it is possible to determine where the hydration is within the DNA binding domain. From the different levels of hydration and the low sequence homology between PU.1 and Ets-1, it was theorized that the primary sequence was responsible for the varying mechanisms of site recognition [66]. Understanding the role of hydration in PU.1 sequence selectivity will provide a clearer understanding of the inherent preference of PU.1 for a given sequence and insight into the mechanisms of sequence selectivity.

2 MAPPING INTERFACIAL HYDRATION IN ETS-FAMILY TRANSCRIPTION FACTOR COMPLEXES WITH DNA: A CHIMERIC APPROACH

 [Nucleic Acids Research, Volume 46, Issue 20, November 2018, Pages 10577-10588, DOI: 10.1093/nar/gky894]

Disclaimer of contributions: Project conception was the work of Gregory M. K. Poon. Data from circular dichroism and osmotic stress experiments were collected and analyzed by Amanda V. Albrecht. Ets-1 and PU.1 high and low-affinity osmotic stress experiments were performed and analyzed by Gregory M. K. Poon. Molecular dynamic simulations were designed, conducted, and analyzed by Gregory M. K. Poon. Cellular reporter assays were performed and analyzed by Hye Mi Kim. Figure design was spearheaded by Gregory M. K. Poon in collaboration with Amanda V. Albrecht and Hye Mi Kim. Manuscript was written in collaboration between Gregory M. K. Poon and Amanda V. Albrecht, with edits performed by Gregory M. K. Poon, Amanda V. Albrecht, and Hye Mi Kim.

2.1 Abstract

Hydration of interfaces is a major determinant of target specificity in protein/DNA interactions. Interfacial hydration is a highly variable feature in DNA recognition by ETS transcription factors and functionally relates to cellular responses to osmotic stress. To understand how hydration is mediated in the conserved ETS/DNA binding interface, secondary structures comprising the DNA contact surface of the strongly hydrated ETS member PU.1 were substituted, one at a time, with corresponding elements from its sparsely hydrated relative Ets-1. The resultant PU.1/Ets-1 chimeras exhibited variably reduced sensitivity to osmotic pressure, indicative of a distributed pattern of interfacial hydration in wildtype PU.1. With the exception of the recognition helix H3, the chimeras retained substantially high affinities. Ets-1 residues could therefore offset

the loss of favorable hydration contributions in PU.1 via low-water interactions, but at the cost of decreased selectivity at base positions flanking the 5'-GGA-3' core consensus. Substitutions within H3 alone, which contacts the core consensus, impaired binding affinity and PU.1 transactivation in accordance with the evolutionary separation of the chimeric residues involved. The combined biophysical, bioinformatics, and functional data therefore supports hydration as an evolved specificity determinant that endows PU.1 with more stringent sequence selection over its ancestral relative Ets-1.

2.2 Introduction

Water strongly mediates the affinity and specificity of protein/DNA interfaces [11, 24]. Hydration contributions to binding arise from differences in molecular interactions available to water in the bulk medium relative to those in the environs of the macromolecules. At protein/DNA interfaces, water may be excluded, interact dynamically with the interface, or form structured bridges that effectively become part of the complex. The influence of interfacial hydration on protein/DNA complexes is therefore complex, but critical for understanding the molecular basis of target selection. For example, restriction endonucleases distinguish sequence-similar DNA substrates based in part on differences in the interfacial hydration of the enzyme/DNA complex [38, 39, 89, 90]. During catalysis, the disposition of water changes as the protein/DNA interface shifts from a substrate-binding conformation to one that is poised for reaction of the transition state [36, 91]. As a result of substrate-dependent hydration contributions, restriction enzymes can discriminate "star" sequences that they will cleave from nonspecific unreactive sequences even though the binding affinities for both DNA are essentially the same [40].

While transcription factors lack catalytic activity, hydration plays similarly important roles in their target DNA recognition. The *E. coli* tryptophan repressor is iconic in this respect [92], and both the crystal structure [54] as well as experiments in solution [41] show that it recognizes operator DNA primarily via water-mediated contacts. In more recent work, we have found that hydration is strongly coupled with site discrimination by the ETS-family protein PU.1, a master transcription factor in hematopoiesis [93]. Typical of this family of transcription factors, PU.1 recognizes cognate DNA sites ~10 bp in length that contain a 5'-GGA(A/T)-3' central consensus. Variants of this motif in which the core consensus is flanked by different sequences bind PU.1 with a broad range of dissociation constants from 10^{-10} to 10^{-7} M under physiologic solution conditions [20]. This *in vitro* selectivity correlates closely with PU.1 cognate sites in native target promoters and enhancers [69].

When probed by osmotic pressure, high-affinity binding by PU.1 is profoundly destabilized due to the accumulation of hydration water in the protein/DNA interface [85]. In addition, binding to low-affinity cognate sites is distinguishable from nonspecific binding in terms of their hydration properties. Low-affinity cognate binding is not osmotically sensitive, while nonspecific binding is strongly stabilized by osmotic pressure, even though the apparent affinities of the two modes differ only marginally in the absence of osmotic stress. Thus hydration distinguishes target recognition by PU.1, which regulates a large genetic network in hematopoietic cells [94], of cognate versus nonspecific binding sites as well as optimal versus low-affinity cognate sites.

The sequence-specific effects of hydration in PU.1/DNA binding prompted us to ask whether they represented a class property of the ETS family, a major lineage of metazoan transcription factors. ETS proteins are united by their DNA-binding domains, known as ETS

domains, which are strongly homologous in structure [67]. They share overlapping DNA preferences that match sequence motifs bound by their full-length parents [70]. Despite these similarities, lineage-restricted ETS proteins such as PU.1 are functionally non-redundant and only partially interchangeable with close phylogenetic relatives *in vivo* [95, 96]. Molecular mechanisms that confer functional specificity to ETS proteins therefore represent a focal point of their molecular biology in development and disease [97]. In particular, general mechanisms that govern the DNA-binding behavior of ETS domains are of fundamental biological interest.

In this connection, we reported that Ets-1, a structural homolog that is co-expressed with PU.1 during hematopoiesis [98, 99], exhibits markedly different DNA-binding properties with respect to interfacial hydration. Although the ETS domains of PU.1 and Ets-1 are divergent in amino acid sequence, they fold into superimposable structures and bind their optimal DNA targets with similarly high affinity as PU.1 in the absence of osmotic stress. However, unlike PU.1, Ets-1 is minimally sensitive to osmotic pressure regardless of DNA sequence [66]. As a functional consequence, we noticed that the two ETS relatives also differ in their levels of site selectivity. Analysis of DNA sequence motifs for the two homologs by information theory shows that PU.1 is significantly more selective in target readout than Ets-1 [100]. The higher stringency by PU.1 is fully maintained in data taken from *in vitro* site selection assays using recombinant ETS domains, as well as ChIP-Seq experiments probing native proteins in human and mouse cells [100]. Thus, PU.1 is more target-selective than Ets-1 *in vivo*, and this stringency is accrued from intrinsic differences in DNA recognition between their ETS domains.

In addition to serving as an experimental probe for studying hydration changes, osmotic stress is a physiologic condition in hematopoietic tissues [101, 102]. We analyzed gene expression data in murine macrophages and found that target genes for PU.1 are significantly over-represented

in osmotically responsive genes [66]. Recently, *hypo*-osmotic stress in K562 leukemic cells was reported to *drive* PU.1 binding to promoter sites in significant excess over Ets-1 [103], in agreement with their osmotic profiles in binding experiments. The differential interfacial hydration in DNA binding therefore also appears to establish the two ETS members' roles in cellular response to physiologic osmotic stress.

The biological significance of hydration in DNA recognition and gene regulation by PU.1 and Ets-1 highlights a need to better understand the physical basis of their hydration differences. Recently, we probed the structure of interfacial hydration in optimal PU.1/DNA complexes by mutating a conserved water-coordinating Tyr residue in PU.1 to Phe [72]. This one-atom difference reduced PU.1 sensitivity to osmotic pressure by ~25% without affecting binding affinity under normo-osmotic conditions. This unexpected observation suggested that hydration of the PU.1/DNA interface might not be essential to binding, and alternative interactions could compensate for reduced hydration contributions to binding affinity.

To resolve the structural basis of hydration in the PU.1/DNA interface and characterize the effects of hydration on binding affinity and selectivity, we used a chimeric approach to map the hydration landscape in the PU.1/DNA interface. Specifically, we took advantage of the strong structural conservation among ETS domains and asked whether the hydration contributions within the DNA contact interface could be mapped to specific secondary structures comprising that surface. We evaluated PU.1 mutants in which each secondary structure comprising the DNA contact surface was replaced one at a time by the corresponding element from Ets-1. Perturbations in the DNA binding affinity of these PU.1/Ets-1 chimeras and their osmotic sensitivity would therefore provide insight into the contribution of the substituted element to interfacial hydration as well as the attendant effects on target affinity and selectivity.

2.3 Materials and Methods

2.3.1 Molecular Cloning

Codon-optimized DNA fragments encoding various PU.1 mutants were synthesized by IDT DNA Technologies (Midland, IA) and subcloned into NcoI/HindIII sites of pET28b vector for overexpression in *E. coli*, or into NheI/HindIII sites of pcDNA3.1(+) for eukaryotic expression. The bacterial constructs harbored additional C-terminal sequences that encoded a cleavage site for thrombin followed by a 6xHis tag. All constructs were verified by Sanger sequencing.

2.3.2 Protein expression and purification

Heterologous overexpression of wildtype (PU.1ΔN167) and chimeric PU.1 ETS domains in BL21(DE3)pLysS *E. coli* was performed as previously described [65]. In brief, expression cultures in LB media were induced at an OD_{600} of 0.6 with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 4 h at 25°C. Harvested cells were lysed by sonication in 0.1 M TrisHCl, pH 7.4 containing 0.5 M NaCl and 5 mM imidazole. After centrifugation, cleared lysate was extracted with Co-NTA and eluted in up to 15 mL of elution buffer containing 150 mM imidazole. The eluate was dialyzed overnight in the presence of 10 U of thrombin (MPBio) against 10 mM NaH_2PO_4/Na_2HPO_4, pH 7.4 containing 0.5 M NaCl, and polished on Sepharose SP (HiTrap, GE). After extensive washing in this buffer, the protein was eluted in a NaCl gradient at ~1 M. Purified protein was dialyzed extensively into various buffers appropriate for the experiments, and diluted as needed with dialysate. Protein concentrations were determined by UV absorption at 280 nm using the extinction coefficients (in M^{-1} cm^{-1}): 22,460. Each construct was verified by MALDI-ToF(+) analysis [**Table 2-1**].

	Molecular weight, Da		
	Expected (E)	*Observed (O)*	%(*O* – *E*)
Wildtype	12,847	12,844	-0.02
Chimeras			
H2	12,843	12,818	-0.2
Loop	12,774	12,784	+0.08
H3	12,894	12,898	+0.03
H3/S3	13,024	13,030	+0.05
S3	12,927	12,957	+0.2
Wing	12,875	12,879	+0.03
N236Y	12,455	12,459	+0.03

Table 2-1 MALDI-ToF analysis of wildtype and chimeric ETS domains of PU.1

Ten ng of purified recombinant protein was examined by MALDI-ToF(+) mass spectroscopy. In all cases, two m/z peaks corresponding to the +1 (base peak) and +2 molecular ions were observed. The expected molecular weights were computed based on elements at natural isotopic abundance and in situ cleavage of the leading methionine by bacterial aminopeptidase [79].

2.3.3 Circular dichroism spectroscopy

Far-UV (190 to 250 nm) spectra were acquired in 10 mM NaH_2PO4/Na_2HPO_4, pH 7.4, and 0.15 M NaCl with a Jasco J-810 instrument.

2.3.4 Osmotic stress experiments

DNA binding experiments by fluorescence anisotropy measurements of a Cy3-labeled DNA probe were performed as described [72, 79, 104]. Briefly, sub-saturating concentrations of wildtype or chimeric PU.1 (10^{-9} to 10^{-8} M) were incubated to equilibrium with graded concentrations of an unlabeled 23-bp DNA duplex oligo harboring the high-affinity PU.1 target site 5'-AGCGGAAGTG-3' or the low-affinity site 5'-AAAGGAATGG-3'. The binding mixtures

contained 10 mM TrisHCl, pH 7.4, 0.15 M NaCl, 0.1 mg/mL BSA, and various concentrations of betaine to achieve the desired osmotic pressure. Solution osmolality was determined by vapor pressure measurements on a Wescor VAPRO 5600 instrument. Steady-state anisotropies $\langle r \rangle$ were measured at 595 nm in 384-well black plates (Corning) in a Molecular Dynamics Paradigm reader with 530 nm excitation. The signal represented the fractional bound DNA probe (F_b), scaled by the limiting anisotropies of the bound $\langle r_1 \rangle$ and unbound states $\langle r_0 \rangle$ as follows:

$$\langle r \rangle = F_b \left(\langle r_1 \rangle - \langle r_0 \rangle \right) + \langle r_0 \rangle \qquad \text{Equation 2-1}$$

F_b as a function of *total* unlabeled DNA concentration was fitted to a mutually exclusive binding model as detailed previously [104] and summarized below for convenience.

2.3.5 Analysis of PU.1/DNA titration data

The effect of osmolytes on DNA recognition, when extrinsically labeled DNA is used as probe, is not accurately reported by direct binding to the probe DNA due to extraneous interactions by the osmolyte with the fluorescent label [105]. We therefore measure the affinity of PU.1 with *unlabeled* DNA indirectly via its inhibition of the PU.1/probe complex. Titration of Cy3-labeled DNA and PU.1 with unlabeled DNA is described by a competitive model in which PU.1 (denoted 01) binds either the labeled probe (1*0; the asterisk denotes probe) or unlabeled competitor (10), but not both, to form the complex 1*1 or 11 respectively. We previously applied this model to describe the competition of PU.1 and other ETS proteins between probe and various unlabeled DNA competitors [66, 79, 106]. The binding polynomial, which is cubic in (11) is,

$$0 = \varphi_0 + \varphi_1[11] + \varphi_2[11]^2 + \varphi_3[11]^3$$

$$\begin{cases} \varphi_0 = -K_{1*1}[10]_t^{\,2}[01]_t \\ \varphi_1 = K_{11}K_{1*1}[10]_t + K_{11}[1*0]_t[10]_t + K_{1*1}[10]_t^{\,2} + 2K_{1*1}[10]_t[01]_t - K_{11}[10]_t[01]_t \\ \varphi_2 = -K_{11}K_{1*1} + K_{11}^{\,2} - K_{11}[1*0]_t - 2K_{1*1}[10]_t + K_{11}[10]_t - K_{1*1}[01]_t + K_{11}[01]_t \\ \varphi_3 = K_{1*1} - K_{11} \end{cases}$$

Equation 2-2

where $K_{1*1} = \frac{[1*0][01]}{[1*1]}$ and $K_{11} = \frac{[10][01]}{[11]}$ are equilibrium dissociation constants for the labeled and unlabeled complex. This mechanistic model therefore directly estimates binding affinities and completely separates any osmolyte-specific effects on probe binding. Note also that the model does not use of the IC_{50} [the so-called Cheng-Prusoff procedure; 107]. $F_b = \frac{[11]}{[11]_t}$ is substituted into Eq. (2-1) in the main text to fit the concentration-dependent anisotropy $\langle r \rangle$. K_{11} (for the unlabeled DNA) is the equilibrium constant of interest reported in the main text. Typically, sub-saturating concentrations of PU.1 (in the absence of competitor) are used to relieve depletion effects that tend to introduce sharp curvature into the data [108]. K_{1*1} can be independently determined from direct titration of the probe with protein alone.

2.3.6 Molecular dynamics simulations

Explicit-solvent simulations were performed with the Amber14SB/parmbsc1 forcefields [109] in the GROMACS 2016.4 environment. Homology models of wildtype and chimeric PU.1ΔN167 in complex with the 23-bp experimental DNA sequence were generated against the co-crystal PU.1/DNA complex [PDB: 1PUE; 77] as a template using the SWISS-MODEL server [110]. The starting models were solvated with TIP3P water and 0.15 M NaCl in a dodecahedral box whose edge was at least 1.0 nm from the closest atom in the model. All simulations were

carried out at an *in silico* temperature and pressure of 298 K (modified Berendsen thermostat) and 1 bar (Parrinello-Rahman ensemble). All bonds were constrained using LINCS. After the structures were energy-minimized by steepest descent, the NVT ensemble was equilibrated at 298 K for 100 ps to thermalize the system followed by another 100 ps of equilibration of the NPT ensemble at 1 bar and 298 K. Production simulation was performed for 400 ns with configurations saved every 10 ps for analysis. Convergence of the trajectories was checked by RMSD from the energy-minimized structures. Average structures were taken as the middle snapshot determined by a Jarvis-Patrick clustering procedure based on the mutual RMSD of the all snapshots from the final 100 ns of the simulations.

2.3.7 Cellular reporter assays

Cellular PU.1 transactivation was measured using a PU.1-dependent EGFP reporter construct as previously described [78]. In brief, the EGFP reporter is under the control of a minimal PU.1-dependent enhancer harboring a 5× tandem of a native cognate site for PU.1. In PU.1-negative HEK293 cells, the reporter was transactivated in the presence of an expression plasmid encoding wildtype full-length PU.1 and a co-translating iRFP marker [83]. 7×10^4 cells were seeded in 24-well plates and co-transfected with a cocktail consisting of the EGFP reporter plasmid (300 ng) and expression plasmids for full-length PU.1 (50 ng) and wildtype/mutant ETS domains (150 ng), using JetPrime reagent (Polyplus, Illkirch, France) according to the manufacturer's instructions. One or both of the expression plasmids were replaced by empty pcDNA3.1(+) vector in negative control samples. Eighteen hours after transfection, the cells were trypsinized and analyzed by flow cytometry using a FCS Fortessa instrument (BD). Live cells were gated for iRFP and EGFP fluorescence using reporter- and full-length PU.1-only controls respectively in FlowJo (BD) before computing the total fluorescence of the dually fluorescent population. Expression of

the ETS domains were quantified with ImageJ software from immunoblots of cell lysate after probing with a polyclonal anti-PU.1 antibody (GenScript A01692) with GAPDH as loading control.

2.3.8 Bioinformatics analysis of sequence motifs

Curated sequence motifs for human PU.1 (SPI1) and Ets-1 derived from ChIP-Seq analysis were culled from the CIS-BP database [111]. Motifs that did not cover the a 10-bp window centered at the 5'-GGA(A/T)-3' consensus were excluded. The position frequency matrix of each motif was analyzed by enoLOGOS [112] to extract the information content (IC) at each position and to generate a sequence logo. No species-related bias in GC content was included in the IC computations (i.e., each base was assumed to be equiprobable).

2.4 Results

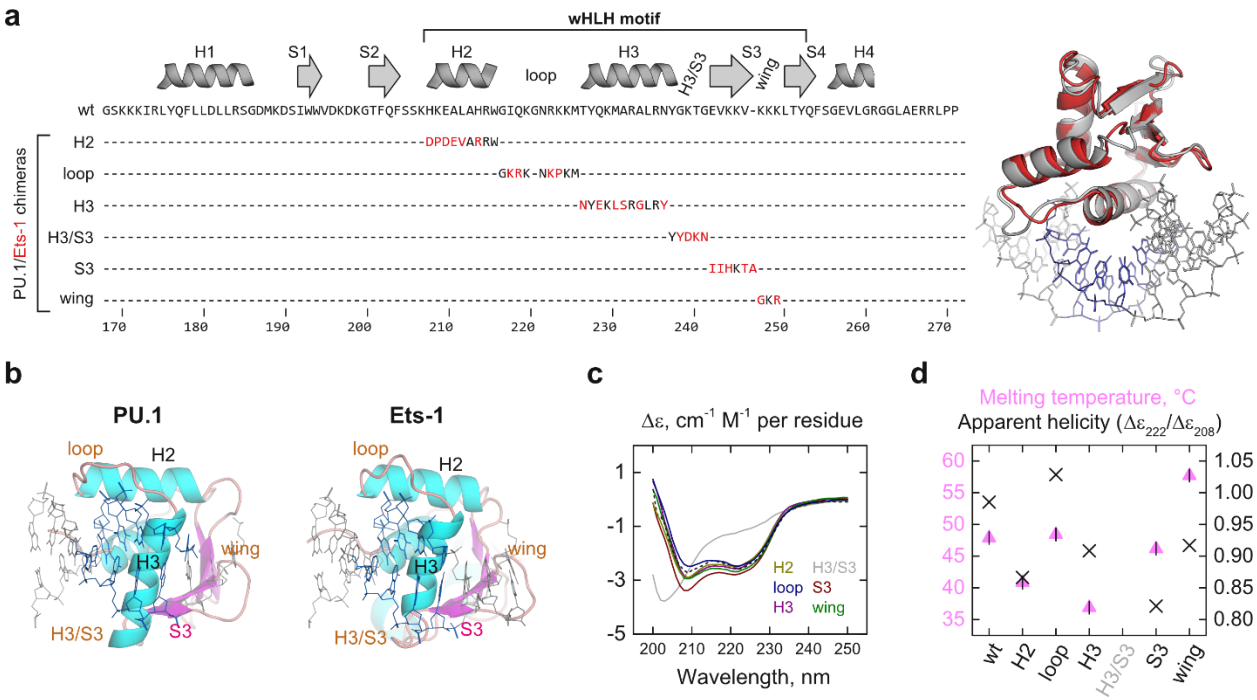


Figure 2-1 A chimeric approach to mapping PU.1/DNA interfacial hydration.

*(**A**) Amino acid sequences of the six PU.1/Ets-1 chimeras. Each of the secondary structure elements that comprise the winged helix-loop-helix DNA-binding motif in PU.1 is replaced with the corresponding element in Ets-1. WT PU.1 and Ets-1 are superimposable in their DNA-bound states (PDB IDs: 1PUE and 1K79). (**B**) Topology of DNA contact interfaces in WT PU.1 and Ets-1. Note the secondary structure assignments of H3/S3 in PU.1 versus Ets-1 as per the crystal structures (PDB IDs: 1PUE and 1K79). (**C**) Far-UV CD spectra of WT and chimeric PU.1 ETS domains at 25°C. (**D**) Melting temperatures of WT and chimeric PU.1 (triangles), plotted with an index for α-helical content taken as the ratio of the molar residue ellipticities at 222/208 nm.*

Our chimeric approach to dissecting the interfacial hydration in the site-specific PU.1/DNA complex is shown in Figure 2-1. The chimeras were designed from a sequence alignment of murine ETS domains [**Figure 2-2**] with a view of maintaining the register of residues that are most conserved in all ETS paralogs. The secondary structural elements that comprise the DNA contact

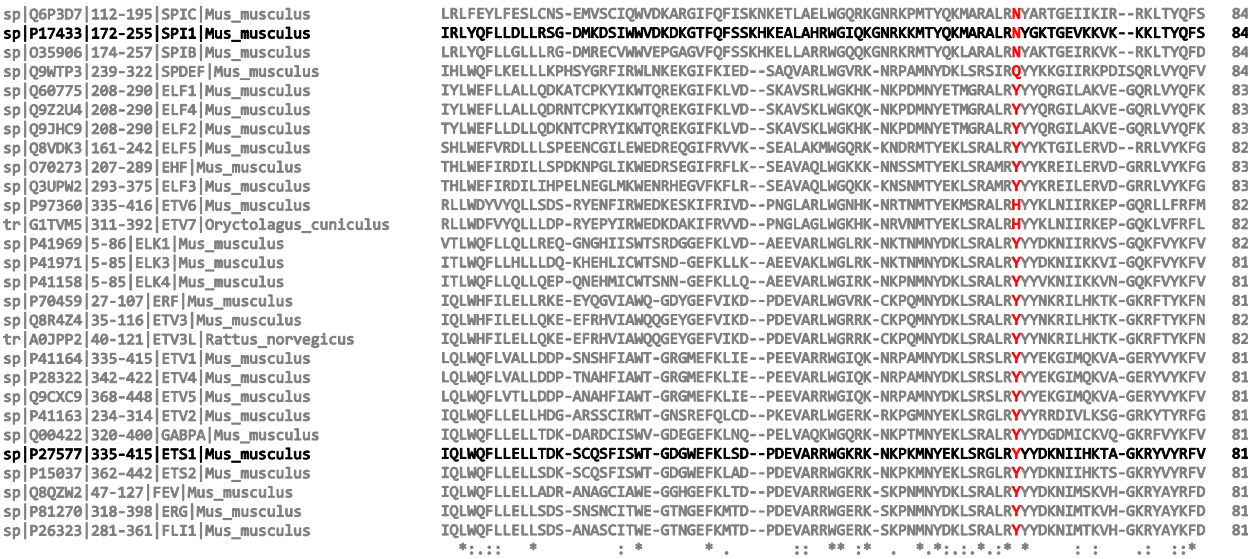


Figure 2-2 Multiple sequence alignment of the murine ETS-family of transcription factors Sequences were culled from UniProt and aligned by Clustal Omega, using the house mouse (M. musculus) as reference organism. In two cases (ETV7 and ETV3L) where an annotated paralog is not available, we used the ortholog from rabbit (O. cuniculus) and rat (R. norvegicus). PU.1 (Spi-1) and Ets-1 are in black. The paralogous residues (Tyr, His, and Gln) corresponding to Asn^{236} in PU.1 are colored in red.

surface in order of primary structure are H2, loop, H3, H3/S3, S3, wing, and S4, the last of which we had previously examined [72] [**Figure 2-1a**]. These elements together constitute the conserved "winged helix-loop-helix" motif [**Figure 2-1b**]. H3 is the groove-binding recognition helix that interacts with the conserved 5'-GGA(A/T)-3' core consensus in ETS binding sites. The extended loop and short wing connect H2 and H3, and S3 and S4 respectively. The H3/S3 segment is a sharp

turn between H3 and S3 that includes several residues assigned as part of H3 in the PU.1/DNA complex [77]. Here, we follow the assignment specified by the Ets-1/DNA co-crystal structure [71].

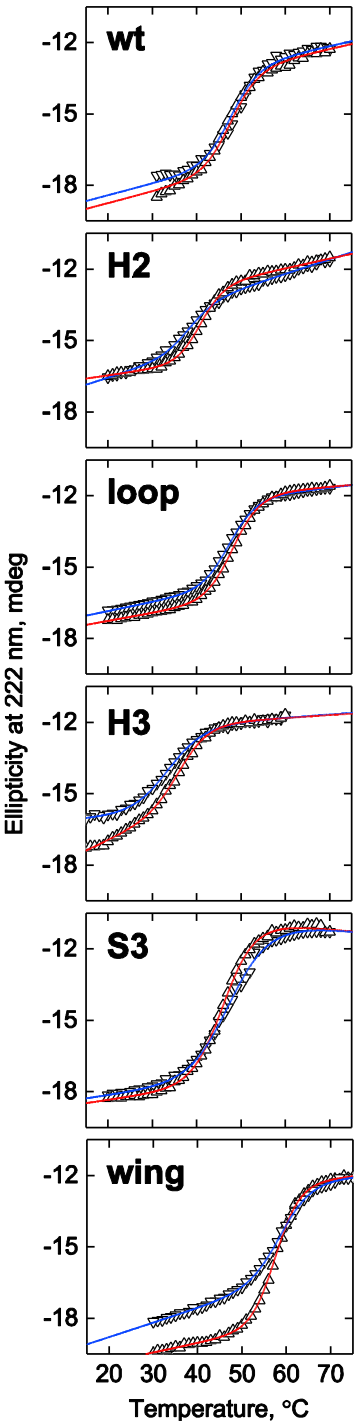


Figure 2-3 Thermal unfolding and refolding of PU.1/Ets-1 chimeras

CD-monitored (at 222 nm) melting (△) and refolding (▽) was carried out with 25 μM wildtype and chimeric PU.1 ETS domains under normo-osmotic conditions at ±45°C/h and a response time of 32 s. The H3/S3 chimera, which did not fold natively as PU.1, was not tested. Curves represent fits to the unfolding/refolding data (red/blue) by a two-state transition [113]. The averaged melting temperature (T_m) of the heating and cooling runs were reported ±1°C in Figure 2d of the main text.

We first addressed whether the chimeras folded correctly relative to wildtype PU.1 by circular dichroism spectroscopy. Five out of the six chimeras exhibited similar secondary structure as wildtype PU.1, with strong α-helical content at 25°C [**Figure 2-1c**]. The lone exception, H3/S3, appeared to be substantially unfolded. Thermal experiments showed that the five well-folded chimeras melted reversibly with melting temperatures (T_m) no more than 10°C different from wildtype [**Figure 2-3**]. The variations in T_m did not correlate with secondary structure contents [**Figure 2-1d**] and were similar to published ETS mutants [72]. Thus, with the exception of H3/S3, the PU.1/Ets-1 chimeras were natively folded under our experimental conditions as wildtype PU.1. These five

chimeras were subsequently interrogated for their DNA-binding properties under osmotic pressure.

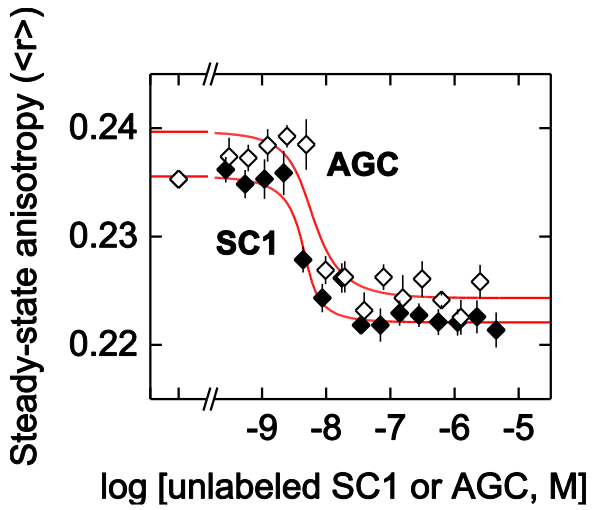


Figure 2-4 The ETS domain of wildtype Ets-1 binds optimal cognate DNA target for PU.1 with similar affinity under physiologic conditions in vitro

The normo-osmotic affinity of the C-terminal ETS domain of wildtype Ets-1 (murine residues 331 to 440, without the auto-inhibitory helices) for its optimal DNA (SC1: 5'-GCCGGAAGTG-3') and optimal DNA for PU.1 (AGC: 5'-AGCGGAAGTG-3') was determined by fluorescence polarization exactly as described in the main text. The affinities for SC1 and AGC are 0.17 ± 0.14 nM and 0.6 ± 0.3 nM, respectively.

To determine the perturbation of the chimeric substitutions on DNA recognition, we measured the binding affinity of the chimeras for an optimal PU.1-binding sequence under osmotically variable conditions. This sequence is also bound by wildtype Ets-1 with high affinity [**Figure 2-4**]. Previously, we have extensively characterized the effect of compatible osmolytes on DNA binding by wildtype PU.1 and Ets-1 and found that they acted quantitatively in a colligative manner (i.e., independent of osmolyte identity), consistent with perturbing specific protein/DNA binding through osmotic pressure [66, 85]. The apparent structural conservation of wildtype PU.1,

Ets-1, and their chimeras (except H3/S3) supports the utility of osmotic stress to probe changes in the latter's interfacial hydration with DNA. Here, we used betaine, a widely used compatible osmolyte [114], to exert osmotic pressure.

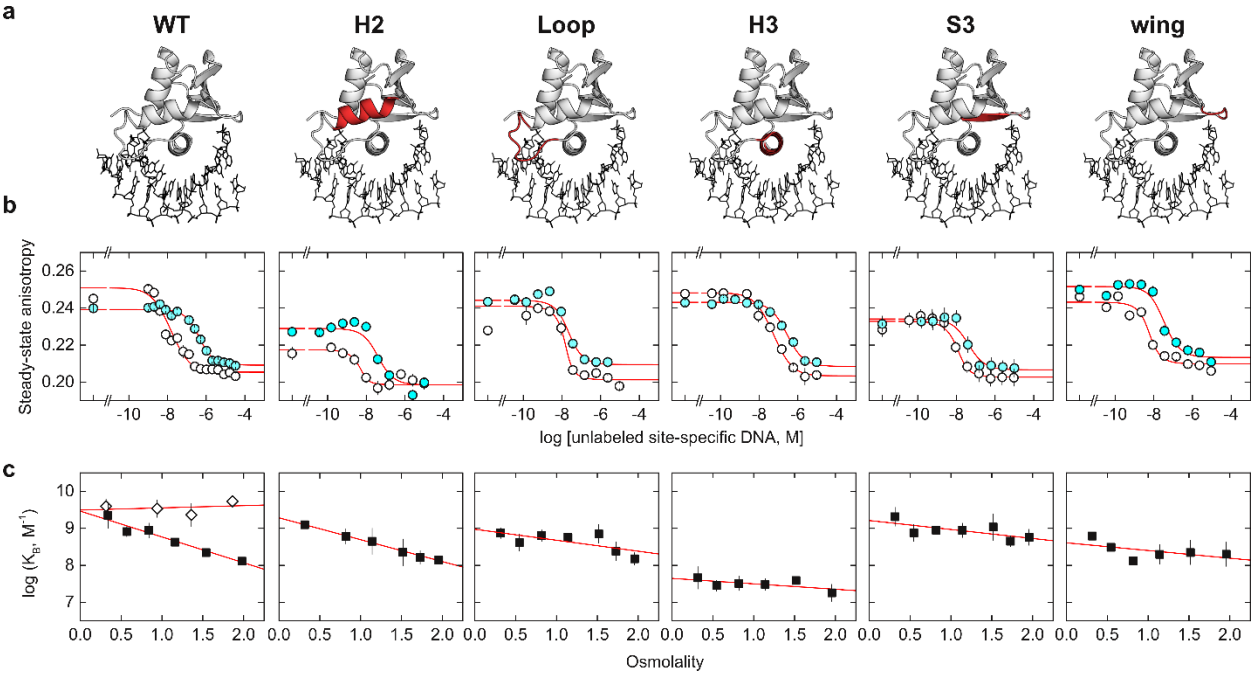


Figure 2-5 PU.1/Ets-1 chimeras reveal a dispersion of osmotic sensitivity and normo-osmotic DNA binding affinities.

(A) Schematic showing the Ets-1 substitutions shaded. (B) Representative competitive titrations under normo-osmotic conditions (0.29 osm, open circles) and at 2 osm (solid). Increasing concentrations of unlabeled-specific DNA competitor displaces protein from a Cy3-labeled DNA probe, lowering its anisotropy. Curves represent fits of Equation (2-1) to the data. Variations in the limiting anisotropies and curvatures in the data arise from differences in affinity of the proteins to the probe and unlabeled DNA, protein concentrations used to acquire a sufficient change in anisotropy, betaine's effects on the photophysical properties of the probe and small variations associated with plate-based detection. These variations are handled by the fitting model and frees the experimenter to select protein concentrations that balance the change in anisotropy with depletion of the titrant, given the wide range of affinities involved. The model

directly estimates the affinity of the protein to the unlabeled target DNA, the parameter of interest (not IC50); see Supplementary Methods for details. (C) Osmotic pressure dependence of the binding affinity. Each point represents the average of three or more replicates ± S.D. WT Ets-1 is included as reference (diamonds). Lines represent linear fits to the data; parametric values of the slope (osmotic sensitivity,) and normo-osmotic affinity (at 0.3 osm) are given in Table 2-1.

By differentially raising the free energy cost of assembling water of hydration in an interface that excludes osmolyte, osmotic pressure imposes a penalty on high-affinity DNA binding by wildtype PU.1. As a result, the logarithm of the binding constant for PU.1 decreases linearly with osmolality by over 1.5 orders of magnitude to over 2 osm above normo-osmotic conditions [58]. This slope may be related to the net uptake or release of water molecules when strongly excluded osmolytes such as betaine are used to exert this pressure [35]:

$$\Gamma \equiv \frac{\partial \log K_{\mathrm{B}}}{\partial\ \mathrm{osm}} = -\frac{\Delta n_{\mathrm{w}}}{55.5 \ln 10} \qquad \text{Equation 2-3}$$

where 55.5 × ln 10 = 128 and a positive value of Δn_{w} indicates net uptake. When probed under identical solution conditions, all PU.1/Ets-1 chimeras exhibited the same linear trend but with variable slopes that were intermediate of wildtype PU.1 and Ets-1 [**Figure 2-5, Table 2-2**]. These shallower slopes indicated that the chimeric complexes retained intermediate levels of interfacial hydration and confirmed these chimeras as functioning hydration probes of the PU.1/DNA interface. In addition to the slope, the DNA-binding affinities at normo-osmotic pressure (0.3 osm) indicated whether the chimeric Ets-1 residues supported high-affinity binding in the PU.1 scaffold.

Together, these two parameters describe the hydration contribution (slope) of a substituted element and whether chimeric Ets-1 residues confer alternative interactions that offset the loss of favorable hydration contributions (affinity at normo-osmotic pressure). Since the elements differ in size and exposure to the bound DNA, we also estimated their per-residue DNA contact surface area based on the wildtype PU.1/DNA complex to facilitate assessment of their binding properties.

	Normo-osmotic affinity $K_D \times 10^{-9}$, M	Osmotic sensitivity Γ, osm^{-1}	Hydration change Δn_w	Contact area with DNA, Å²/residue
Wildtype	0.34 ± 0.06	-0.70 ± 0.05	90 ± 6	21 ± 1 (overall)
Chimeras				
H2	0.80 ± 0.05	-0.59 ± 0.01	76 ± 1	4.0 ± 0.1
Loop	1.3 ± 0.4	-0.30 ± 0.14	39 ± 18	23 ± 1
H3	22 ± 10	-0.15 ± 0.09	19 ± 11	37 ± 1
H3/S3	nd	nd	Nd	3.7 ± 0.2
S3	0.49 ± 0.12	-0.24 ± 0.10	31 ± 13	9.7 ± 0.2
Wing	1.6 ± 0.1	-0.21 ± 0.15	27 ± 19	30 ± 1

Table 2-2 Binding properties of WT and chimeric PU.1 ETS domains

Site-specific affinity of WT and chimeric ETS domains of PU.1 was measured by competition by unlabeled target DNA against a Cy3-labeled DNA probe for protein. Representative titrations are shown in Figure 2-5B. Steady-state fluorescence anisotropy data were fitted with Equation (2-1). Averages ± SE of three to five independent replicates were used to determine the osmotic pressure dependence of the binding affinity as shown in Figure 2C. Normo-osmotic affinity is given as the equilibrium dissociation constant KD (reciprocal binding constant). Osmotic sensitivity (Γ) is the linearly fitted slope of the osmotic pressure dependence. The implied change in hydration water molecules was determined from via Equation (2-3). nd, not determined. The

DNA contact area was computed from the WT PU.1/DNA complex (PDB ID: 1PUE) as the difference in solvent-accessible surface area of the structure in the presence and absence of the bound DNA (averaged for the two asymmetric units in the model).

The reduction in osmotic sensitivity among the chimeras tracked roughly with the per-residue DNA-contact area of the substituted element in wildtype PU.1 (Table 2-2). Thus, the H2 chimera, which is the least exposed to the DNA among the well-formed chimeras, also exhibited the smallest hydration perturbation among the tested chimeras. Conversely, the H3 and wing chimeras, which represent the most contacted elements in the interface, yielded the greatest losses in osmotic sensitivity. Chimeric substitution was more perturbative than expected at S3 based on DNA contact surface area, suggesting additional hydration contributions by this element above the others. Since destabilization by osmotic pressure reflects the favorable hydration contributions to binding, the data indicated that hydration was not localized to any one or small subset of secondary structures comprising the DNA contact surface in wildtype PU.1.

Binding affinity under normo-osmotic conditions also tracked roughly with osmotic sensitivity and DNA-contact area, but the perturbations were generally smaller in magnitude. Although chimeric substitutions of the H2, loop, S3, and wing of Ets-1 into a PU.1 significantly blunted sensitivity to osmotic pressure, binding affinity was reduced by no more than 5-fold at normo-osmotic pressure. This was not a significant change in affinity given the >400-fold span of affinities with which wildtype PU.1 binds cognate DNA sites [20]. Thus, favorable hydration contributions in wildtype PU.1 to affinity were offset by alternative interactions by the Ets-1 residues in these chimeras. In the case of the loop, S3, and wing chimeras, the decoupling of

hydration from high-affinity binding rendered their binding profiles similar to Ets-1, which binds optimal DNA with similar affinity as PU.1 but minimal osmotic sensitivity (c.f. Figure 2-5a).

2.4.1 The H3 chimera is significantly impaired in both normo-osmotic binding and osmotic sensitivity

The H3 chimera differed from the other constructs in its significant loss of binding affinity (>50-fold) as well as osmotic sensitivity. This was a biologically significant level of reduction, approaching within ~10-fold of the affinity of wildtype PU.1 for low-affinity specific sites [20]. Thus, unlike the other structural elements, Ets-1 residues in the recognition helix H3 were incompatible with high-affinity binding in a PU.1 scaffold.

To identify non-conserved residues responsible for the low tolerance to chimeric substitution in H3, we noticed that this helix is the most conserved element among ETS domains. Every H3 position is either fully or conservatively substituted, except residue 236. In wildtype PU.1, this residue is Asn, which is unique to PU.1 and its two closest ETS relatives (Figure 2-2). In the wildtype PU.1/DNA complex, Asn^{236} is exclusively engaged in water-mediated contacts with the 5'-GGA-3' core consensus in the major groove. In contrast, the corresponding position in wildtype Ets-1 is Tyr^{395}, which is engaged in a direct H-bond with the exocyclic N6 of 5'-GG<u>A</u>-3' in the consensus [**Figure 2-6a**]. To probe the significance of Asn^{236} in PU.1, we mutated it to Tyr. The resultant N236Y chimera was conformationally conserved with wildtype PU.1 [**Figure 2-6b**], indicating that this substitution was non-perturbative in the unbound protein. Nevertheless, binding by N236Y was identical to the H3 chimera [**Figure 2-6c**]. Chimeric substitution of Asn^{236} alone was therefore sufficient to reproduce the binding profile of the full H3 chimera.

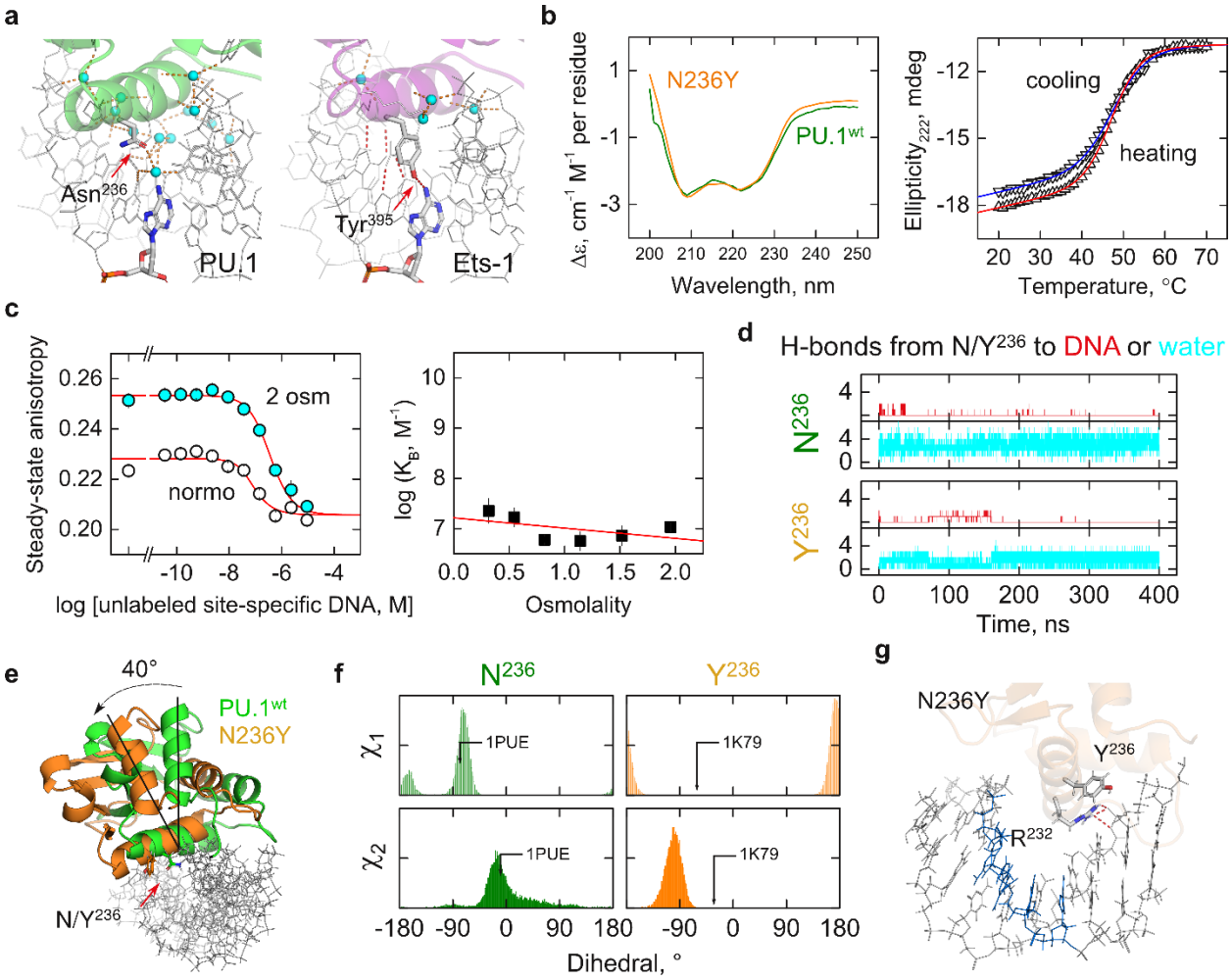


Figure 2-6 Asn[236] in the recognition helix H3 is a keystone residue in DNA binding by PU.1.

(A) Co-crystal structures of WT PU.1 and Ets-1 show contrasting roles for Asn[236] in PU.1 and its Tyr[395] counterpart in Ets-1. (B) The point chimera N236Y showed an identical CD spectrum at 25◦C and reversible unfolding as WT PU.1. (C) Decompensated DNA binding by N236Y with loss of osmotic sensitivity. (D) All-atom MD simulation of WT and N236Y bound to the experimental DNA target. H-bonds (within 3.5 Å, ±30°) between Asn[236] or Tyr[236] and DNA or water molecules show an initial burst of direct DNA contacts by N236Y and partial desolvation of Tyr[236] between 50 and 150 ns. (E) Alignment of the average structure from the final 100 ns shows a 40° re-orientation of the chimera relative to the DNA. (F) Distribution of the two

standard sidechain dihedrals of Asn^{236} and Tyr^{236} in the final 100 ns. Arrows mark values from the co-crystal structures in Panel A. (G) Mapped average configurations of Tyr^{236} and Arg^{232} in H3 of N236Y, the latter of which has lost its absolutely conserved consensus contact (blue).

To understand the loss of high-affinity binding by N236Y in greater detail, we performed all-atom molecular dynamics (MD) simulations of the site-specific wildtype PU.1 and N236Y complexes using the co-crystal wildtype structure as template. Following ~0.2 μs of equilibration in unconstrained production, the wildtype complex achieved a stable ensemble around the canonical ETS/DNA conformation. Integrity of the interfacial hydration in the wildtype complex was evaluated in terms of contact between Asn^{236} and DNA versus water. The strong hydration of Asn^{236} in preference over direct DNA contact demonstrated consistency of the MD results with experimental data on wildtype PU.1 [**Figure 2-6d**]. In contrast, the N236Y chimera exhibited altogether different dynamics. During the equilibration period (from ~50 to 150 ns), it made frequent direct contact with N6 of 5'-GGAA-3' with attendant partial desolvation of the sidechain. Strikingly, it then underwent a major transition that abolished these direct DNA contacts. The average final structures (from a clustering procedure) revealed a large re-orientation of the chimeric protein with respect to the target DNA [**Figure 2-6e**]. More precisely, the protein pivoted ~40° near the midpoint of H3 in the major groove but without significant rearrangement of its tertiary structure. While the wildtype Asn^{236} sidechain ensemble effectively maintained the same dihedrals as in the co-crystal structure, the Tyr^{236} sidechain in N236Y had swung out of position in the new orientation and no longer pointed into the major groove [**Figure 2-6f**]. Moreover, canonical interactions between Arg^{232} with nucleobases in the consensus, a conserved H3 contact

engaged by all specific ETS/DNA complexes (including PU.1), were replaced by contacts with the DNA backbone just beyond the core consensus [**Figure 2-6g**].

2.4.2 Chimeric perturbation at Asn^{236} is evolutionarily sensitive

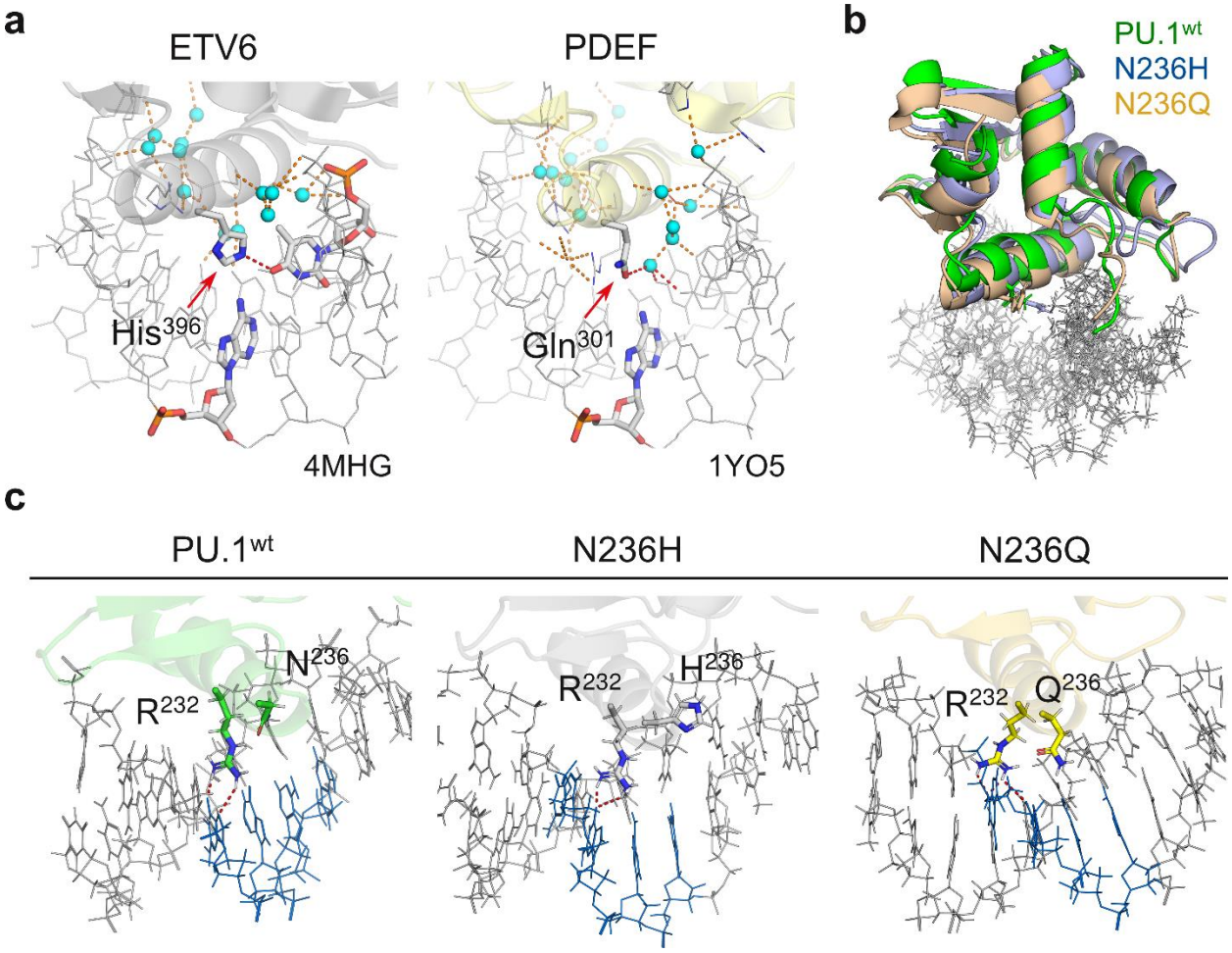


Figure 2-7 Evolutionarily conservative chimeras at Asn^{236} are also structurally conservative.

(A) The ETS domains of ETV6 and PDEF are phylogenetic intermediates between PU.1 and Ets-1 with similar levels of co-crystallographic interfacial hydration as PU.1. Their counterparts of Asn^{236} in PU.1 are marked by arrows. (B) All-atom simulated complexes of N236H and N236Q with the experimental DNA target maintained the same canonical configuration as WT PU.1. As

shown, the average equilibrated structures from the final 100 ns were aligned by the bound DNA only to demonstrate their configurational homology. (C) Average equilibrated structures of simulated complexes show the conservation of direct contacts with the 5 -GGAA-3 consensus by the sidechain of Arg^{232}.

In addition to Tyr, Asn^{236} in PU.1 are replaced by His and Gln in phylogenetic intermediates between PU.1 and Ets-1, namely ETV6, ETV7, and PDEF (Figure 2-2). In the co-crystal structure of ETV6 [115], the corresponding His^{396} makes a direct H-bond not with 5'-GGAA-3' but O4 of 5'-TTCC-3' in the opposite strand of the core consensus [**Figure 2-7a**], with a slightly lower amount of crystallographic hydration as PU.1. In the co-crystal structure of PDEF [116], Gln^{311} makes only water-mediated contacts similarly as Asn^{236} in PU.1. Since interfacial crystallographic hydration is correlated with evolutionary separation among ETS domains [64], we asked whether substitution of Asn^{236} in PU.1 with His or Gln would be less perturbing than Tyr found in Ets-1. To address this question, we simulated DNA-bound N236H and N236Q chimeras under identical *in silico* conditions as the other complexes. Unlike N236Y, both chimeric complexes maintained the canonical complex configuration of wildtype PU.1 [**Figure 2-7b**] and crucially preserved direct contacts with core consensus DNA by Arg^{232} [**Figure 2-7c**; c.f. Figure 2-7g]. Nevertheless, examination of the sidechain dihedrals showed that His^{236} had flipped out of position similarly as Tyr^{236} in N236Y, with a correspondingly similar hydration profile [**Figure 2-8**]. In N236Q, the dihedral ensemble for Gln^{236} adopted configurations that resembled those observed for Gln^{301} in the PDEF co-crystal structure. Thus, chimeric substitutions at position 236 of PU.1 with evolutionarily proximal residues were more compatible with conserving the

canonical ETS/DNA complex than a more ancestral counterpart, namely Tyr found in ETS paralogs such as Ets-1.

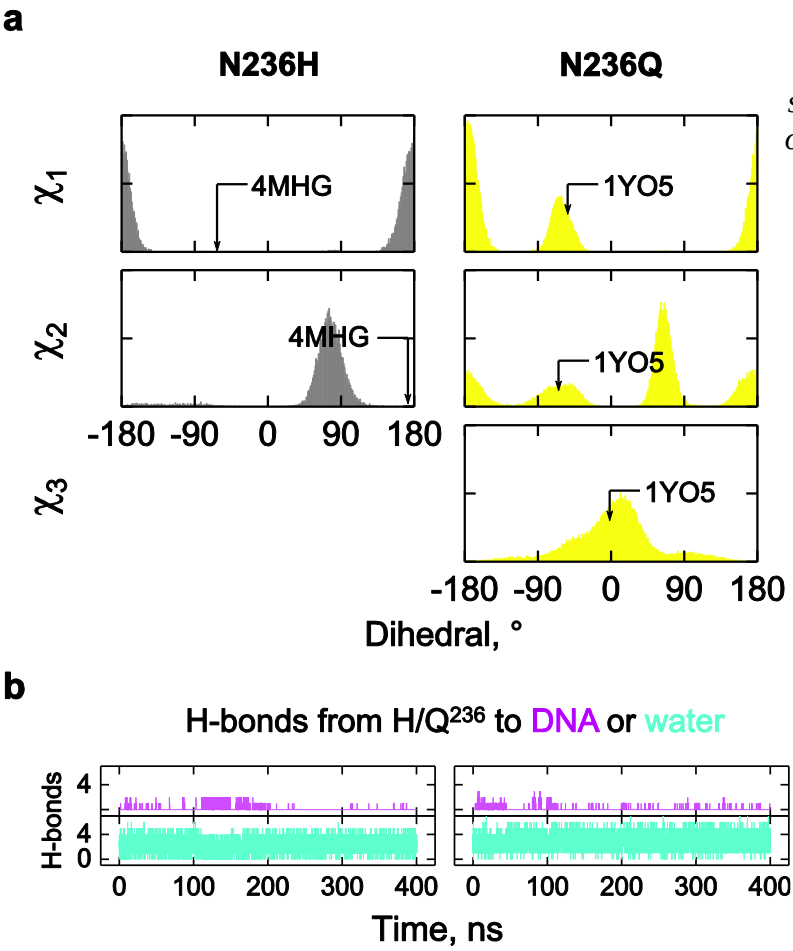


Figure 2-8 Standard sidechain dihedrals and H-bonding of the chimeric residue in simulated N236H and N236Q mutants of the PU.1 ETS domain

***A**, Equilibrated trajectories from the final 100 ns of simulation were analyzed as for Figure 4f in the main text. Arrows mark the dihedrals from the co-crystal structures of ETV6 [4MHG; 115] and SPDEF [1YO5; 117], two proximal ETS relatives that harbor His and Gln at the corresponding positions as Asn^{236} in wildtype PU.1. **B**, H-bond contacts (<3.5 Å and ±30° between heavy donor/acceptor pairs) formed by the chimeric residue with DNA and water. Note the similar burst of direct DNA contacts formed by His^{236} in N236H between 100 to 200 ns as Tyr^{236} in N236Y (Figure 4d in the main text), but is absent in wildtype or N236Q.*

To probe the functional significance of evolutionary variations of Asn^{236}, we tested the N236Y/H/Q mutants as dominant-negative inhibitors of PU.1 transactivation in cells. We co-transfected HEK293 cells with expression plasmids encoding full-length PU.1 and wildtype or

mutant ETS domain, together with a PU.1-dependent reporter plasmid [**Figure 2-9a**]. The reporter expressed EGFP under the control of a synthetic enhancer consisting of a pentameric PU.1-binding λB motif from the $Ig\lambda_{2\text{-}4}$ enhancer [78]. The full-length PU.1 transgene was cloned with an iRFP construct via a co-translating 2A peptide to enable isolation of PU.1-expressing cells by flow cytometry [83]. In PU.1-negative HEK293 cells, the reporter was negligibly activated by endogenous transcription factors in control transfectants lacking ectopic full-length PU.1 [**Figure 2-9b**]. PU.1-dependent EGFP fluorescence was measured in the subpopulation of iRFP-expressing cells (i.e., upper-right quadrant). Co-transfection of constant amounts of full-length and ETS-encoding genes showed that the N236Y mutant inhibited full-length PU.1 ineffectively relative to wildtype ETS domain. The N236H and N236Q mutants were both more effective inhibitors than N236Y with the N236Q being statistically similar to wildtype [$p = 0.05$; **Figure 2-9c**]. In summary, functional studies in cells confirmed MD and biophysical characterizations that mutation of Asn^{236} perturbed DNA binding by PU.1 in an evolutionarily sensitive fashion.

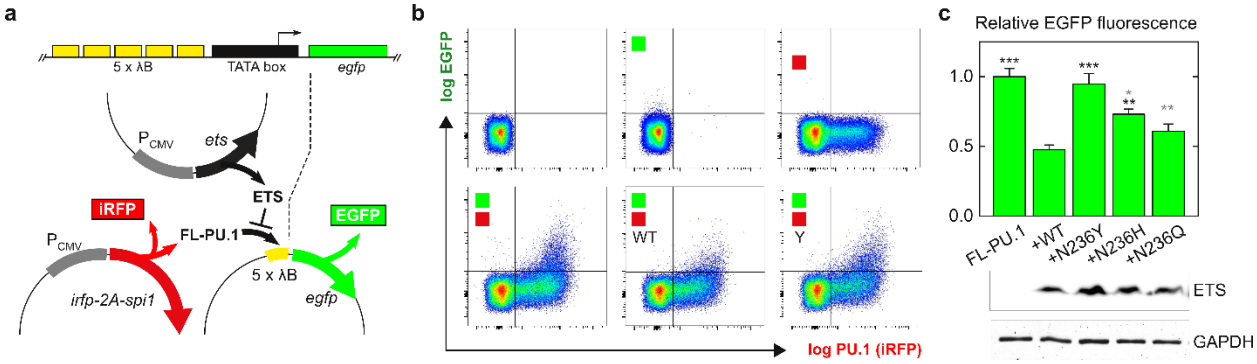


Figure 2-9 Inhibition of PU.1 transactivation by evolutionarily conservative H3 chimeras. Cognate DNA binding by PU.1 ETS mutants N236X (X = Y, H or Q) was evaluated functionally as dominant-negative inhibitors of PU.1 transactivation in HEK293 cells. (A) Schematic of a PU.1-dependent EGFP reporter under the control of an enhancer consisting of a 5 × tandem λB motif. Cells expressing ectopic full-length (FL) PU.1 was gated via a co-translating iRFP marker. (B) Representative flow cytometric data on transient HEK293 transfectants and

*controls. Colored squares denote co-transfection with the FL-PU.1 expression plasmid or EGFP reporter plasmid. WT and Y denote co-transfection with expression plasmid for the WT or N236Y mutant of the PU.1 ETS domain. See 'Materials and Methods' section for details. Axes represent logarithmic intensities of the iRFP marker and EGFP reporters. (C) EGFP fluorescence, adjusted for the expression of the ETS domains (by immunoblot) and normalized relative to the intensity of the no-ETS sample as mean ± S.D. (N = 3), and analyzed by one-way ANOVA with Tukey's post-hoc tests. *,**,***: $p < 0.05, 0.005, 5 \times 10{-}4$ (black, versus WT ETS; gray, versus N236Y).*

2.4.3 Interfacial hydration confers sequence selectivity to PU.1

Although ETS proteins bind a spectrum of cognate DNA variants harboring a 5'-GGA(A/T)-3', they differ in selectivity for these variants. One general approach to parameterizing sequence selectivity is to apply information theory to bound sequence motifs [118]. Specifically, the information content (IC) of a sequence motif quantifies the sequence discriminating power of the ligand in terms of the number of binary bits, ranging from 0 per base position if all four bases are equally probable, to 2 if the position is fully specified by a single base [119]. For PU.1 and Ets-1, we have previously determined the IC of sequence motifs from *in vitro* selection experiments as well as ChIP-Seq data on genomic binding [100]. As the representative DNA logos [120] in **Figure 2-10a** show, PU.1 exerts higher selectivity at the two 5' distally flanking bases (positions -3, -2) and four 3' flanking bases (positions +3 to +6) over Ets-1. Relaxation of selectivity at the core +3 position from A to A/T by a N236Y mutation has been previously reported [121]. The remaining base positions at which PU.1 is more selective are proximal to the wing, S3, and loop whose chimeras are strongly affinity-compensated despite significant losses of

osmotic sensitivity (Table 1). An updated compilation of *in vivo* data show that over their full 10-bp binding sites (out of a maximum IC of 2 × 10 = 20 bits), PU.1 exhibits an IC that is over ~3.5 bits or 30% higher than Ets-1 [**Figure 2-10b**]. This margin is even more striking on account of the 6 bits of IC fixed by the invariant core consensus at positions 0 to +2.

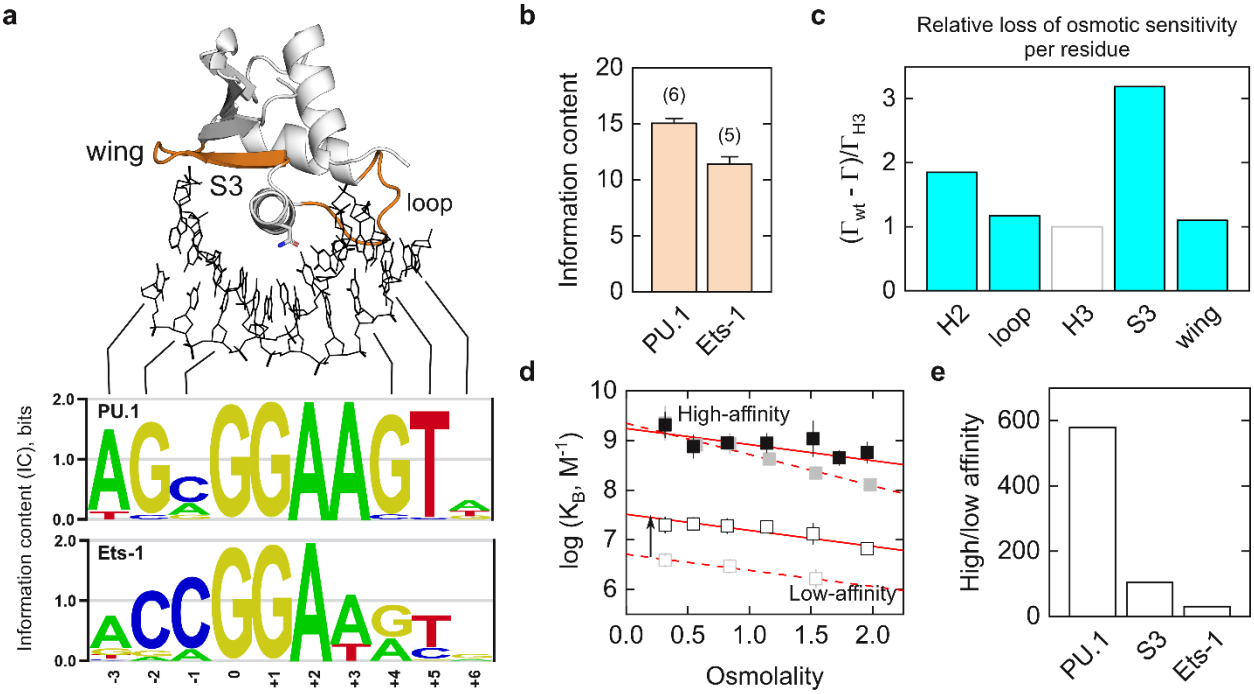


Figure 2-10 Osmotic sensitivity is associated with stringent DNA sequence discrimination by PU.1.

(A) Positions in PU.1-binding sites with higher stringency over Ets-1 are closely contacted by structural elements that exhibit strong affinity compensation upon loss of osmotic sensitivity, colored in orange using the PU.1/DNA co-crystal structure as model. Asn236 is shown as sticks. The two in vivo sequence motifs are taken from Jolma et al. (48). (B) Average ICs ± SE of in vivo bound sequence motifs (10 bp) curated at the CIS-BP database (see Supplementary Methods), with the number of motifs in parenthesis. (C) Per-residue loss of osmotic sensitivity among the chimeras relative to WT, normalized to the H3 chimera. Only the four chimeras colored in blue exhibit affinity compensation. (D) Osmotic pressure dependence of high- (solid symbols) and low-affinity

binding (open) by the S3 chimera (black) relative to WT PU.1 (gray; c.f. Figure 2A). (E) Ratio of affinities for high- versus low-affinity targets by the S3 chimera in comparison with WT PU.1 and Ets-1.

The positional correlation of bases at which PU.1 is differentially selective with affinity-compensated DNA-contacting elements led us to ask whether the compensatory interactions by chimeric Ets-1 residues would be less sequence-selective than the native contacts that they replaced. To test this hypothesis, we compared the affinity of the S3 chimera, which exhibited the greatest *per-residue* loss in osmotic sensitivity [**Figure 2-10c**] and strongest affinity compensation for its optimal DNA target over an established low-affinity PU.1 binding site [20]. The S3 chimera bound low-affinity DNA ~5 fold more strongly than wildtype PU.1 [arrow in **Figure 2-10d**]. Since both wildtype and S3-chimeric PU.1 bound high-affinity DNA equally well, the ratio of high/low affinity binding by S3 chimera was ~5-fold lower than wildtype PU.1 [**Figure 2-10e**], and ~3-fold higher than wildtype Ets-1 for its respective high- and low-affinity sequences [76]. In addition, whereas the osmotic sensitivity of wildtype PU.1 is sequence-dependent, the S3 chimera exhibited similar osmotic sensitivity for both DNA targets, indicating a lack of hydration contribution to selectivity. Bioinformatics analysis and direct measurements therefore showed that introduction of Ets-1 residues into PU.1 reduced osmotic sensitivity and its attendant sequence selectivity.

2.5 Discussion

With the exception of H3/S3, PU.1/Ets-1 chimeras of the DNA-contact surface retain the conserved structural framework characteristic of ETS domains. The chimeric perturbations on osmotic sensitivity therefore reveal the contribution of each structural element to hydrating DNA-binding interface. The chimeras exhibit blunted osmotic sensitivity relative to wildtype PU.1

roughly in step with their per-residue contact area with DNA. Overall, the data shows that interfacial hydration in the PU.1/DNA interface is not locally mediated by any one or small subset of elements and suggests a network of distributed interactions involving the contact surface as a whole. The S3 chimera is a notable standout in that it contributes on a per-residue basis more hydration interactions than the other DNA-contact elements.

Although all chimeras lose osmotic sensitivity, only the H3 chimera also exhibit significant loss of binding affinity in the absence of osmotic stress. The normo-osmotic affinities of the H2, loop, S3, and wing chimeras are within 5-fold of wildtype PU.1, which is not a significant reduction given the over 400-fold span in affinity between high- and low-affinity cognate binding [20]. In previous work [72], we reported that disruption of a crystallographic water contact by the interfacial mutation Y252F in PU.1 reduced osmotic sensitivity by ~25%, but without effect on normo-osmotic binding affinity to the same high-affinity DNA used here. The chimeric data for H2, loop, S3, and wing therefore reinforces and generalizes the concept of affinity compensation. As osmotic stress raises the free energy cost of accumulating interfacial water, the decoupling of osmotic sensitivity and binding affinity indicates that the loss of favorable hydration contributions in these chimeras can be offset by Ets-1 residues to achieve almost the same affinity. The compatibility of alternative interactions with high-affinity binding over large portions of the DNA-contact surface reflects the low primary structure conservation of these domains among paralogous ETS proteins.

An important functional consequence of hydration in DNA recognition by PU.1 is that relative to a sparsely hydrated relative Ets-1, the incorporation of water molecules increases the sequence selectivity of binding. The sequence motifs bound by the two native ETS paralogs show deficits in information content for Ets-1 at most base positions flanking the 5’-GGA(A/T)-3’

central consensus. These flanking bases are most closely contacted by DNA-contacting elements whose Ets-1 chimeras exhibit the highest reduction of osmotic sensitivity with affinity compensation. As demonstrated by the S3 chimera, substitution with Ets-1 residues results in a strongly affinity-compensated species with increased affinity for a low-affinity target, narrowing the selectivity gap between optimal and low-affinity DNA. Moreover, the S3 chimera shows no hydration contribution to selectivity, in contrast to the strong sequence dependence in osmotic sensitivity found in wildtype PU.1. Taken together, the data supports the hypothesis that divergent primary structures of PU.1 and Ets-1 are linked to greater target selectivity for PU.1.

2.5.1 Significance of the H3 recognition helix in the ETS motif

Unlike the other well-folded chimeras, H3 exhibits significant loss of normo-osmotic binding as well as sensitivity to osmotic pressure. The PU.1-specific H3 residues are therefore highly adapted to the other elements of the PU.1 ETS domain. Of all the secondary structures in the ETS domain, H3 exposes the largest fraction of its surface area to the DNA contact interface. As shown by the N236Y mutant, which recapitulates the binding properties of the full H3 chimera, the canonical ETS/DNA complex configuration is disrupted by a large re-orientation of the protein relative to the target DNA site. In agreement with this structural prediction, functional assays using a PU.1-dependent EGFP reporter confirmed that the N236Y chimera exhibited significantly impaired DNA binding in live cells. H3 is also the most conserved element in the ETS domain (Figure 2-2). Residue 236 is the only H3 position that exhibits significant variation: Asn, Tyr, His, Gln. On the one hand, all four residues are related by exactly a single transition in their codons (A↔G or C↔T) and align with the overall phylogenetic separation among ETS paralogs [70]. On the other hand, the relative compatibility of these residues with the PU.1 scaffold in both MD and reporter assays correspond to their relative hydrophobicity in the order Tyr > His > Gln > Asn

at near-neutral pH [122]. Thus, the significance of H3 in PU.1/DNA binding is also closely tied to interfacial hydration and suggests an evolutionary transition in the essential recognition helix towards the acquisition of interfacial hydration.

2.5.2 The structure and thermodynamics of interfacial hydration in ETS/DNA complexes

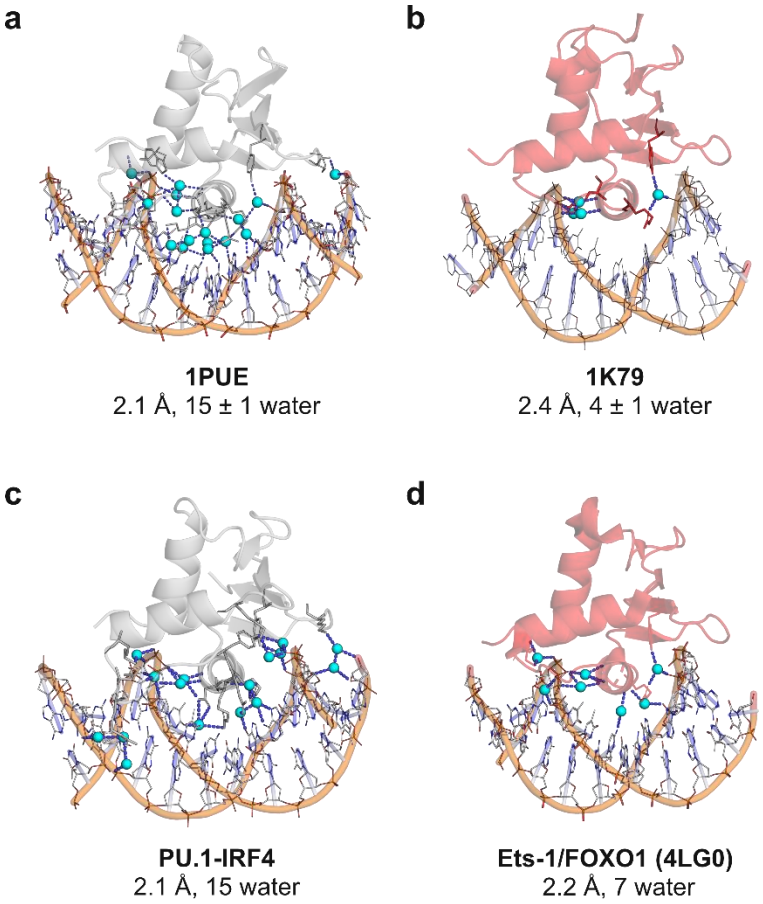


Figure 2-11 Crystallographic interfacial water in PU.1 and Ets-1 complexes with high-affinity DNA

Only bridging waters within 3.5 Å between donor/acceptor are enumerated; non-bridging crystallographic hydration (whose disposition in the unbound states cannot be inferred only from the complex) is excluded. The refinement resolutions are given. ***a and b****, Binary ETS/DNA structures of the PU.1 (1PUE) and Ets-1 (1K79). These co-crystals contain two asymmetric units per unit cell. The bridging water in both asymmetric units is counted and averaged.* ***c and d****, PU.1 and Ets-1 in ternary structures with a binding partner: IRF4 for PU.1 [123], FOXO1 for Ets-1. The binding partner is not rendered. These structures have only one asymmetric unit per unit cell.*

To date, structures of site-specific DNA complexes for approximately half of the 28 human and mouse ETS domains have been solved crystallographically. At comparable resolutions and irrespective of binding partners, PU.1/DNA complexes harbor more than twice as many bridging water molecules in its interface relative to Ets-1/DNA complexes [**Figure 2-11**]. Many of the corresponding residues in Ets-1 make direct DNA contacts instead. The excess bridging water in PU.1 is found all along its DNA interface, consistent with the distributed pattern reported by osmotic stress. In addition to ordered bridging water, more dynamic and weakly held water that is nevertheless energetically important do not appear in the structures [124]. The global hydration change due to DNA binding is detected thermodynamically as the osmotic pressure dependence of the binding affinity via Eq. 2-3. High-affinity binding by wildtype PU.1 is associated with the net uptake of ~90 water molecules, while Ets-1 binding is net hydration-neutral (c.f. Figure 2-5a). From the PU.1/DNA co-crystal structure, the total water-accessible surface area of the protein/DNA interface is 985 ± 24 Å² (average of the two asymmetric units in 1PUE). Taking the nominal cross-sectional area of a water molecule as 9 Å², the osmotic sensitivity of wildtype PU.1 implies that up to $\frac{90 \text{ water} \times 9 \ {}^{\text{Å}^2}/_{\text{water}}}{985 \text{ Å}^2} = \sim 80\%$ of the its interface with DNA becomes net hydrated. Comparison of this total hydration change with the co-crystal structure therefore identifies distinct populations of hydration water. Beyond the ~15 ordered bridging hydration as captured crystallographically, an additional number of more dynamic water molecules are involved in hydrating the PU.1/DNA complex. Correspondingly, a quantity of weakly held water is net displaced upon formation of the Ets-1/DNA complex.

In addition to affinity, which corresponds to the free energy change, the hydration contributions to DNA binding by PU.1 and Ets-1 are manifest in the underlying thermodynamic parameters. High-affinity PU.1 binding exhibits an unusually small negative change in heat

capacity [32, 84], a parameter for which large magnitudes are interpreted as dehydration (release of hydration water) upon binding [125]. Subsequent calorimetric comparisons of ETS domains from wildtype PU.1 and Ets-1 show that binding is entropically more favorable for Ets-1 than PU.1 and accompanied by a more negative heat capacity change [66]. These thermodynamic signatures

are consistent with a configurational penalty paid by dynamically restricted water in the PU.1/DNA interface. Osmotically blunted but affinity-compensated PU.1/Ets-1 chimeras are therefore expected to be more entropically driven with more negative changes in heat capacity, effectively tending toward the thermodynamic profile of wildtype Ets-1.

2.6 Conclusion

The ETS-family transcription factors PU.1 and Ets-1 bind their respective cognate DNA targets with similarly high affinity under normo-osmotic conditions but exhibit profound differences in their sensitivity to osmotic pressure. By interrogating chimeric ETS domains consisting of targeted substitutions of Ets-1 residues into PU.1, the present data squarely implicate interfacial hydration as a specificity determinant in DNA recognition by PU.1. More generally, the relationship between affinity and specificity continues to be a matter of fundamental interest in protein/DNA interactions. The structurally conserved ETS factors offer a biological venue in which this question can be addressed, while also providing insight into the likely functional roles of these specific transcription factors. Given the prominent role of ETS domains in directing the molecular properties of the full-length proteins, this study also supports chimeras as a rational approach to dissect other functional and evolutionary relationships of ETS transcription factors.

3 DISSECTING HYDRATION AND DNA DYNAMICS FROM DIRECT READOUT VIA MODIFIED NUCLEOBASES

This work is pre-submission and is currently in the writing phase.

3.1 Introduction

All identified life forms require water for survival, which is vital for function down to the molecular level. Water acts to stabilize various structures such as DNA, RNA, and proteins [11, 24, 42]. Cells contain a complex environment made of solutes, DNA, RNA, proteins, and water. A proper balance is key to function and, therefore, the life of the cell. Water found throughout the cellular structures is essential for various functions beyond folding and stabilizing structures and can play critical roles in modulating reactions [33]. To better understand the role water plays in moderating cellular events, it is possible to study water activity by osmotic stress. Previously, osmotic pressure has been utilized to understand water's role in the cellular membrane, DNA, proteins, and protein-DNA complexes [33]. To understand the functionality of fundamental structures such as proteins, we start in an *in vitro* setting. The protein of interest is produced and grown to be studied in isolation, allowing for a greater understanding of the essential function before expanding to a living organism.

Waters of hydration act as markers for specific sequence recognition in protein-DNA complexes [31, 40]. Previously, thermodynamic studies of protein-DNA interactions have shown that hydration around the protein and DNA is released upon the complex's formation due to steric complementarity and the loss of solvent-accessible surface area [11, 28, 126]. Historically, it was understood that there is little room for water in the protein-DNA interface in site-specific complexes, such as with the BamHI specific DNA complex [127]. When BamHI is bound specifically to DNA, the protein and DNA make direct hydrogen bonds, releasing the waters that

previously covered the binding domain surface; when bound to a nonspecific DNA, the complex has 120-150 more waters than the specific complex [39]. However, other protein-DNA complexes have been shown to uptake waters upon binding specific sequences rather than actively excluding them, such as PU.1 from the ETS (E26 transformation-specific) family of transcription factors. It was previously shown that PU.1 is osmotically sensitive in a sequence-dependent manner. The binding affinity to specific DNA is decreased 10-fold at 0.5 osmolal, proving that PU.1 binds DNA specifically through water-mediated contacts and that by removing these contacts, the complex is destabilized [64, 72, 128]. To stabilize protein-DNA complexes, various well-studied interactions are utilized, including direct hydrogen bonds, van der Waals interactions, electrostatic interactions, and hydrophobic contacts. In contrast, water-mediated hydrogen bonds remain poorly explained [11].

DNA forms a double helix consisting of major and minor grooves. In the grooves are located hydrogen bonding functional groups from the DNA bases; these functional groups form a hydrogen-bonding pattern that signals to the protein that this stretch of DNA is a specific sequence for the protein of interest [11]. It is understood that proteins recognize their target sites by either base or shape readout mechanisms. The base readout mechanism has direct interactions between the protein and DNA of interest. In shape readout, the global and local shape of DNA is essential for a proper fit of the protein to form the protein-DNA complex [1]. However, these models do not consider water-mediated hydrogen bonds between the amino acids and DNA and do not account for changes within the protein on a localized level. A previous study with the EcoRI endonuclease-DNA complex, which makes 12 hydrogen bond contacts in the co-crystal structure, displayed that the hydrogen bonding pattern was crucial for specific binding [11]. Nevertheless, other protein-DNA complexes have been discovered that do not contain any direct

hydrogen bond interactions, like the trp repressor-operator-specific complex, which binds solely via water-mediated contacts [11]. Previous studies have established that various proteins bind DNA specifically and nonspecifically by multiple mechanisms and interactions. However, disparities exist in understanding how the protein differentiates between specific and nonspecific sequences that diverge by as little as a single base.

To better understand how proteins differentiate between specific sequences, we have chosen to study the ETS family of transcription factors. This family is ideal for studying sequence specificity as there are 28 homologs of ETS transcription factors in the human genome, all of which bind to a cognate DNA sequence containing 5'-GGAA-3'. ETS proteins can bind DNA with a range of affinities (~100 fold differences) as determined by the flanking regions around the cognate site [20, 65, 66, 79]. ETS family members PU.1 and Ets-1 are characterized by profoundly different biophysical mechanisms when binding high, low, and nonspecific DNA sequences in terms of hydration, conformational dynamics, electrostatics, discrimination between cognate sequences, and methylation [66, 79, 100, 128]. However, it is not well understood how the conserved ETS domain distinguishes between DNA sequences. To elucidate the mechanism of how PU.1 differentiates between DNA sequences, a series of DNAs were designed which alter the direct and indirect readout pattern of the cognate DNA sequence 5'-GGAA-3'. The DNAs were paired with wild-type (WT) PU.1 and the PU.1 chimera S3, in which the secondary structure that contacts the flanking region of DNA in the ETS domain was swapped with the corresponding region from Ets-1 [128]. The comparison of PU.1 vs. S3 provides the means to determine the role of hydration in conjunction with the altered hydrogen-bonding pattern. Previously, we showed that the interfacial hydration network iconic to PU.1 is a mechanism for sequence selectivity but not a requirement for high-affinity DNA interactions [72, 128]. Studying

the DNA mutations with PU.1 and S3 by osmotic stress alongside constructing a thermodynamic profile will enable us to parse the mechanism of how PU.1 determines which sequence to bind specifically.

We hypothesize that when binding DNA, PU.1 utilizes both direct and indirect readout in combination with a hydration network to form sequence-specific complexes and that removing any of these aspects individually will decrease sequence selectivity. The hydration pattern of PU.1 is utilized as a method for stabilizing the structure under less than ideal situations, allows PU.1 to be more sequence selective, and makes it less sensitive to methylation over its ancestral relative Ets-1, making PU.1 a pioneer transcription factor [64, 73, 79]. Therefore, we use changes in hydration to study how the specific functional groups on DNA affect the mechanism of binding. In high-affinity complexes, there is a net uptake of water, while PU.1/low-affinity DNA is insensitive to changes in hydration, and PU.1/nonspecific DNA is characterized by a release of water upon binding [73]. From this, we can examine how PU.1 binds to DNA when we make point modifications to the functional groups on the DNA bases and access the effect on the uptake/release of water. Pairing these modified DNA sequences with the S3 chimera provides a model to understand how the hydration network works to compensate for any changes in binding affinity.

3.2 Materials and Methods

3.2.1 Molecular Cloning

PU.1 chimeric constructs were previously designed and examined to map out the location of hydration in the PU.1 DNA binding domain [128]. In short, the PU.1 chimeras were created by utilizing codon-optimized DNA fragments of specific sequences in PU.1 that make up secondary structures in the DNA binding domain in which particular sections were replaced with

the corresponding sequence from Ets-1. The DNA was synthesized by IDT DNA Technologies (Midland, IA, USA) and was subcloned into pET28b vectors with the cut sites NcoI/HindIII and harbored a c-terminal cleavage site for thrombin before a 6xHIs tag. All constructs were verified by Sanger sequencing.

3.2.2 Protein expression and purification

Purification of PU.1 (PU.1ΔN167) and PU.1 ETS domain chimera, S3, followed the procedure as previously described [128]. All protein constructs were heterologously overexpressed in BL21(DE3) *E. coli*; constructs were induced by adding 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6 for 4 hours at 25°C. After collecting the cells, the sample was sonicated in 0.1M Tris-HCl, pH 7.4 with 0.5 M NaCl, and 5 mM imidazole, followed by centrifugation. The clear lysate was collected and purified with Co-NTA and eluted with elution buffer containing 150 mM imidazole in up to 15 mL of volume. The eluted sample was then dialyzed overnight in buffer containing 10 mM NaH_2PO_4/Na_2HPO_4, pH 7.4 containing 0.5 M NaCl with 10 U of thrombin (MPBio). The samples were then purified for a final time on Sepharose SP (HiTrap, GE) and eluted after equilibrating in the buffer in a NaCl gradient at ~1 M. The final protein product was dialyzed overnight into the buffers required for the following experiments. UV absorption at 280 nm was utilized to determine protein concentrations.

3.2.3 DNA constructs

DNA constructs were designed utilizing the high-affinity PU.1 DNA sequence 5'-AGCG**G**AAGTG-3' as a template where only the second guanosine of the consensus sequence GGAA was exchanged with a modified nucleotide. The modified nucleobases utilized consist of inosine (dI), 2-aminopurine (d2-AP), and Iso-deoxyguanosine (Iso-dG)/Iso-deoxy-5-methylated-cytosine (Iso-$d^{5m}C$). The complimentary reverse sequence remains identical to that of high-

affinity DNA for inosine and 2-aminopurine; however, 2-aminopurine was also examined while bound to cytosine as well as thymine. Iso-deoxyguanosine is paired opposite an Iso-deoxy-5-methylated-cytosine. A control sequence was also analyzed consisting of the forward high-affinity strand paired with a reverse strand containing a 5-methylated-cytosine ($d^{5m}C$). The modified DNA strands were synthesized by IDT DNA Technologies (Midland, IA, USA) and annealed in a 1:1 molar ratio in the buffers required for the following experiments.

3.2.4 Osmotic stress experiments

DNA binding experiments were performed as previously described, except with different DNA sequences [72, 79, 104, 128]. Fluorescence anisotropy measurements of a Cy3-labeled DNA probe were performed to measure the DNA-binding affinity of PU.1 and S3 individually to various DNA sequences either alone or with the neutral osmolytes betaine or acetamide. PU.1 or S3 was incubated at sub-saturating concentrations (10^{-9} to 10^{-8}), with increasing unlabeled 23-bp DNA duplex oligo concentrations. The 23-bp DNA sequences utilized contained variations of the known PU.1 high-affinity sequence 5'-AGCGGAAGTG-3', in which the second guanosine base of the consensus G**G**AA was exchanged with the non-natural bases inosine (dI), 2-aminopurine (d2-AP), and iso-deoxyguanosine (Iso-dG). The protein, labeled, and unlabeled DNA sequences were incubated to equilibrium in a binding mixture containing 10 mM TrisHCL, pH 7.4, 0.15 M NaCl, 0.1 mg/mL bovine serum albumin, and a variety of concentrations of the selected osmolyte to reach the desired osmotic pressures. The equilibrated samples were measured by steady-state anisotropy (*r*) at 595 nm in 384-well black plates (Corning) in a Molecular Dynamics Paradigm reader with excitation at 530 nm. The signal represents the fractional bound DNA probe (Fb), which was scaled by the limiting anisotropies of the bound (r1) and unbound states (r0):

$$(r) = F_b((r_1) - (r_0)) + (r_0)$$ Equation 3-1

The Fractional bound DNA probe (F_b) is a function of the total concentration of unlabeled DNA and was fitted to a mutually exclusive binding model, as previously described [104, 128].

3.2.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry experiments were performed as previously described [66]. In short, DNA and purified protein were co-dialyzed in 10 mM NaH_2PO_4/Na_2HPO_4 at a pH of 7.4 with 150 mM NaCl. A set volume and concentration of DNA were placed into the well, and protein was titrated into the DNA sample to perform the titration. The titration was performed on a VP-ITC MiocroCalorimeter, and the data was plotted using Origin Software. PU.1 was titrated into the DNA at sub saturating concentrations and increased until a molar ration of above 2 was reached. The enthalpy change was determined for both the 1:1 and 2:1 complexes.

3.3 Results

3.3.1 Modified DNA bases provide a means of targeting direct and indirect readout individually and simultaneously.

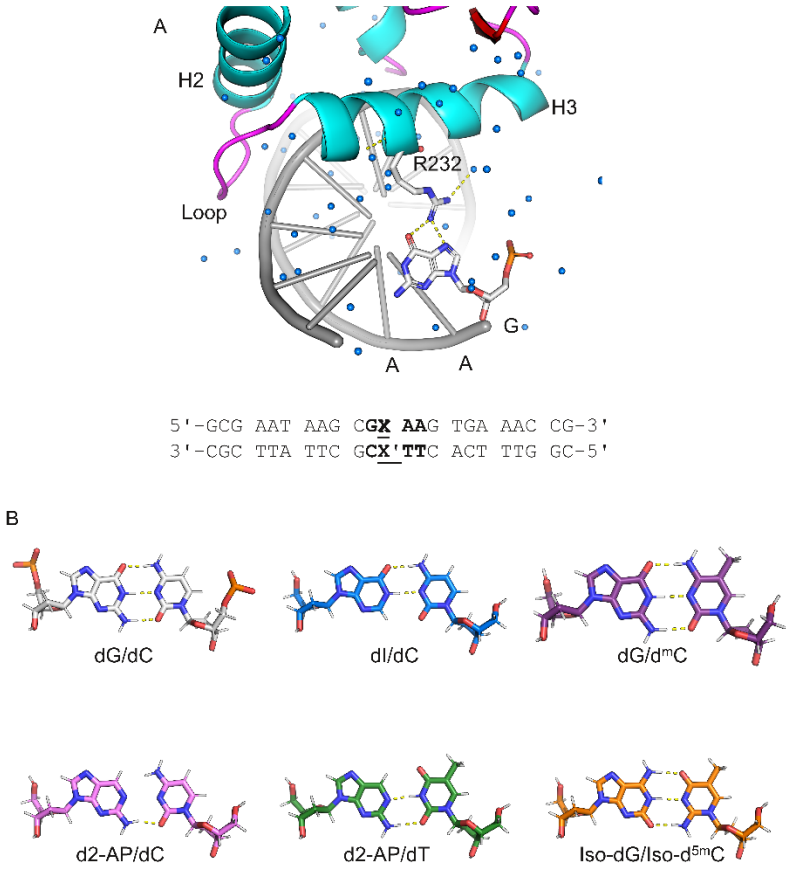


Figure 3-1: Method for altering the pattern of hydrogen bond donors and acceptors in the PU.1 core consensus sequence of high-affinity DNA. A) PU.1 bound to the core consensus sequence 5'-GGAA-3', highlighting the direct hydrogen bond interactions between guanosine and Arg 232 in the recognition helix H3. This interaction was explicitly targeted due to it being a conserved interaction throughout the ETS family. B) The modified DNA bases are shown here, with a Watson and Crick guanosine cytosine pairing for comparison. An inosine (dI) base was chosen as this modified base has lost the minor groove NH_2, providing a means of examining the indirect readout mechanism of PU.1 for high-affinity DNA. 2-Aminopurine (d2-AP) provides a means of examining the role of direct readout by removing the major groove carbonyl. 2-Aminopurine was paired with both cytosine and thymine individually. Iso-deoxyguanosine (Iso-dG) and Iso-deoxy-5-methylated-cytosine (Iso-$d^{5m}C$) were chosen to retain the proper guanosine cytosine base pairing while swapping the hydrogen bond donor and acceptor pattern of the major and minor

grooves. Guanosine bound to 5 methylated cytosine ($d^{5m}C$) was examined to determine if the results from Iso-dG were due to the swapped position rather than the added methylation on Iso-dC.

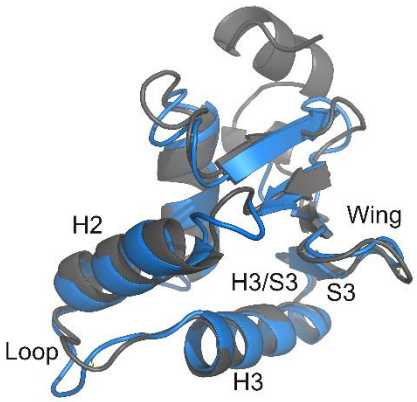
The DNA binding domain of PU.1 is susceptible to alterations in DNA sequences- changes in the core consensus sequence lead to nonspecific binding, and changes in the flanking regions can significantly affect binding affinity leading to low-affinity interactions. To understand the role direct and indirect readout has in PU.1/DNA binding, a series of modified DNA sequences were designed in which direct interactions between the DNA and protein, and interactions between the forward and complementary DNA strands were removed or altered. These specific modifications make it possible to determine the role of direct and indirect readout in protein-DNA specificity. Our approach to altering the high-affinity PU.1 DNA sequence is shown in Figure 3-1. Utilizing the core consensus sequence 5'-GGAA-3' required for high-affinity interactions in the ETS family, the second G of G**G**AA was exchanged with specific deoxyguanosine derivatives. The second guanosine of the consensus sequence was explicitly chosen due to the conserved direct interaction between the carbonyl on position 6 of guanosine and arginine 232 (Arg 232) of helix 3 (H3) of the DNA binding domain; this direct interaction is conserved in all ETS family members (Figure 3-1A).

The deoxyguanosine derivatives were explicitly chosen to simultaneously alter either direct readout, indirect readout, or direct and indirect readout. The deoxyguanosine derivatives examined here are inosine (dI), 2-aminopurine (d2-AP), and iso-deoxyguanosine (Iso-dG) paired with iso-deoxy-5-methylated cytosine (Iso-$d^{5m}C$). While designing these DNA sequences containing deoxyguanosine derivatives, it was essential to retain the B form DNA structure as

much as possible while removing or exchanging specific functional groups. The majority of the derivatives were paired with cytosine to maintain the same complementary DNA sequence as the PU.1 high-affinity sequence. However, two derivatives were paired with a modified cytosine or thymine to retain the proper Watson-Crick base.

Pairing dI with a cytosine provides a path for probing the role of the minor groove in indirect readout through the loss of the amino group distal to the protein/DNA interface. Likewise, the direct readout can be assessed by removing the carbonyl with d2-AP to result in the loss of the conserved direct hydrogen bond between the recognition helix (H3) and DNA. Unlike the other derivatives explored here, d2-AP was paired with both cytosine and thymine individually due to the d2-AP/dC forming a wobble pair [129-131]. In contrast, d2-AP/dT retains two of the three hydrogen bonds characteristic of the dG/dC base pair and maintains Watson-Crick-like pairing [129]. Examining d2-AP/dT provided a means to explore whether the results from the d2-AP/dC were due to the loss of the direct hydrogen bond between the carbonyl and Arg 232 or due to the wobble pairing. The d2-AP construct provides a means of altering direct readout between the protein and DNA while not affecting indirect readout. Iso-dG was paired with Iso-d^{5m}C to retain the three hydrogen bonds between the G/C base pair while swapping the positions of the amino group and the carbonyl, thus altering both direct and indirect readout of the DNA. It was not possible to obtain a non-methylated Iso-dC. Hence, as a control, we examined the effects of guanosine bound to a 5-methylated-cytosine (dG/d^{5m}C) to determine whether the results of the Iso-dG/Iso-d^{5m}C were due to the position exchange, the methylation, or a combination of the two.

3.3.2 Chimeric constructs provide a pathway for determining the level of hydration in protein/DNA interactions.



	H2	Loop	H3	H3/S3	S3	Wing
PU.1	HKEALAHRW	GIQKGNRKKM	TYQKMARALRN	YGKTG	EVKKY-	KKK
Ets-1	DPDEVARRW	GKRK-NKPKM	NYEKLSRGLRY	YYDKN	IIHKTA	GKR

Figure 3-2 Mapping the series of chimeric constructs on the DNA binding domain of PU.1.

A series of chimeric constructs were previously designed in which the structures of the DNA binding domain (DBD) of PU.1 were individually exchanged with the corresponding structures from Ets-1. Here is a modified schematic displaying the series of chimeras that were designed [128]. PU.1 is shown in blue and Ets-1 in grey. The structures chosen to swap all make contact with DNA, whether directly or via water-mediated interactions. The chimeras examined previously include the H2, Loop, H3, H3/S3, S3, and Wing chimeras [128]. Below the structure is the primary amino acid sequence of both PU.1 and Ets-1. The residues that make up PU.1 are colored in blue, and the residues corresponding to Ets-1 are in black. The residues that are retained between the two sequences are colored blue.

Previously, we designed and examined a series of chimeric constructs [128]. The secondary structure elements of the DNA binding domain (DBD) of PU.1 were individually swapped with the corresponding sequence in the ETS family member Ets-1 (Figure 3-2) [128]. This method was possible due to the highly conserved structure of the DBD of the ETS family, known as the ETS domain (Figure 3-2A). In the prior study, we determined the regions that

impacted the hydration network of PU.1 the most. We found that the beta-sheet S3 significantly decreased osmotic sensitivity while retaining high-affinity level binding [128]. The recognition helix H3 had the most significant decrease in binding affinity, with hydration matching low-affinity interactions. The other chimeras displayed various osmotic sensitivities and binding affinities that fell between the extremes of S3 and H3.

Here, we are focusing on the S3 chimera in comparison to PU.1. The S3 chimera was chosen as it retains the highest binding affinity and the lowest hydration level compared with the rest of the chimeras examined previously [128].

3.3.3 PU.1 is highly sensitive to the chemical substituents of guanosine in high-affinity DNA, with guanosine derivatives resulting in near low-affinity levels of binding and hydration.

To test how PU.1 is affected by the changes to guanosine in the high-affinity sequence, a series of fluorescence anisotropy experiments were performed in which PU.1 was paired with the high-affinity (HA), low-affinity (LA), and modified DNA sequences (Figure 3-3, top row). The fluorescence anisotropy measurements were performed starting with sub-saturating concentrations of the protein of interest and the fluorescently labeled DNA to avoid the 2:1 complex. A non-fluorescently labeled DNA strand was then titrated into the sample in increasing concentrations to provide competition between the labeled and non-labeled sequences. The change in anisotropy was measured as the protein exchanged positions from the fluorescently labeled strand to the non-labeled strand. The shift in anisotropy was then calculated as the equilibrium dissociation constant (K_D) of the protein leaving the labeled strand and binding the non-labeled strand. Increasing osmotic pressure allows for the determination of the effect

hydration has on binding affinity. Since PU.1 utilizes a hydration network to form the high-affinity protein/DNA complex, examining PU.1 bound to the modified sequences under increasing osmotic pressure provides insight into the role of hydration in PU. 1's affinity for DNA at the core consensus sequence.

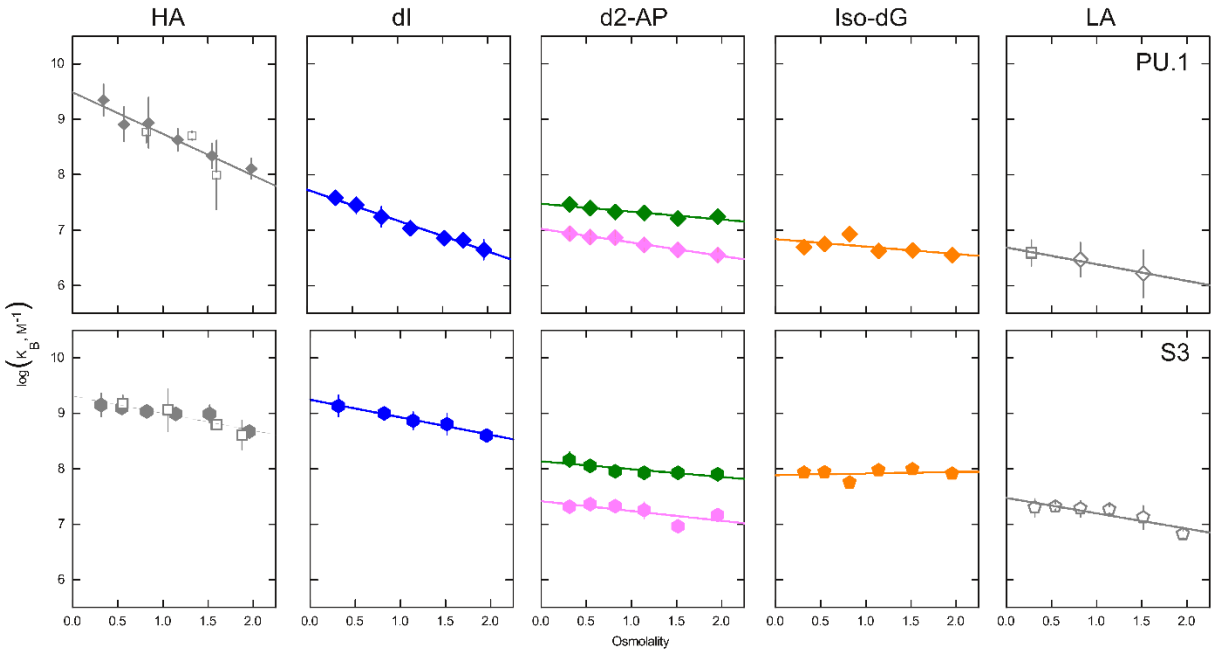


Figure 3-3: PU.1 and the S3 chimera display differing responses to modified DNA, with S3 displaying consistently higher binding affinities under normo-osmotic conditions than PU.1. Fluorescence anisotropy binding experiments performed with increasing osmotic pressure (Betaine in solid colors and open for LA, acetamide in open squares) provide a means of quantifying the role of hydration in protein/DNA interactions. Starting from the top row and going from left to right, PU.1 was paired with six different DNA sequences: high-affinity (HA solid grey), inosine (dI, blue), 2-aminopurine bound with cytosine (d2-AP/dC, pink), 2-aminopurine bound with thymine (d2-AP/dT, green), iso-deoxyguanosine bound to iso-deoxy-5-methyl-cytosine (Iso-dG/Iso-d^{5M}C, orange), and low-affinity (grey open). PU.1 was unable to bind to the modified DNAs as strongly as with the high-affinity sequence, as seen by the decreased binding affinity under normo-osmotic pressure. However, the binding affinity

progressed by the level of disruption to the consensus sequence. The I DNA was the least disruptive to binding affinity, and d2-AP/dC had the greatest effect on affinity, most likely due to the wobble pair. Binding affinity is slightly recaptured in PU.1 bound to d2-AP/dT, which reinstates a second hydrogen bond, thus stabilizing the interaction between d2-AP and dT. The modified bases also drastically decrease hydration to levels comparable to the low-affinity complex, except dI. The PU.1/dI complex displays similar hydration levels to that of the high-affinity complex, suggesting that the loss of the amino group in the minor groove on guanosine is not a critical component of the hydration network of PU.1. The lower row shows S3 bound to the same mutated DNA sequences as PU.1 but with differing results. The S3 chimera is less sensitive to changes in the DNA bases and binds the mutated DNA sequences consistently stronger than PU.1. S3 is much less sensitive to the changes in binding affinity when bound to dI DNA, as seen by the HA level binding and the drastically decreased level of osmotic stress. S3 continues to bind the d2-AP constructs and the Iso-dG/Iso-d^{5M}C stronger when compared to PU.1, with similar patterns of hydration.

In Figure 3-3, a comparison is made of the change in dissociation constant vs. the difference in osmotic pressure. Starting at the top row, PU.1 is bound to HA DNA that retains ~90 waters upon forming the complex. The PU.1/HA complex is destabilized with an increase in the osmotic pressure, as indicated by the negative slope [128]. To calculate the change in

hydration from the slope of the binding curves, equation 3-2 is utilized where a negative Δn_w indicates a net release of hydration.

$$\Gamma \equiv \frac{\partial \log K_B}{\partial osm} = -\frac{\Delta n_w}{55.5 \ln 10} \qquad \text{Equation 3-2}$$

The characteristic binding of PU.1 to HA DNA provides the baseline for native PU.1 interactions and underscores the importance of the hydration network in the high-affinity complex [128]. PU.1 was also examined with LA DNA, resulting in a drastic decrease in binding affinity and hydration. The PU.1/LA complex retains only ~38 waters, and binding affinity is reduced ~100 fold [128]. Both the HA and LA bindings will be utilized in comparison to the modified DNA sequences.

		normo-osmotic affinity (K_d, M)	***osmotic sensitivity (Γ, osm^{-1})***	***Hydration change (Δn_w)***
PU.1	**HA**	**0.34 (± 0.06) x 10^{-9}**	**-0.70 (± 0.05)**	**90 ± 6**
	dI	**2.75 (± 0.45) x 10^{-8}**	**-0.56 (± 0.03)**	**72 ± 4**
	d2-AP/dC	**1.18 (± 0.12) x 10^{-7}**	**-0.24 (± 0.02)**	**31 ± 3**
	d2-AP/dT	**3.43 (± 0.33) x 10^{-8}**	**-0.14 (± 0.03)**	**18 ± 4**
	Iso-dG/Iso-d^{5m}C	**2.03 (± 0.32) x 10^{-7}**	**-0.13 (± 0.08)**	**17 ± 10**
	LA	**2.59 (± 1.41) x 10^{-7}**	**-0.30 (± 0.04)**	**38 ± 5**
S3	**HA**	**0.49 (± 0.12) x 10^{-9}**	**-0.24 (± 0.10)**	**31 ± 13**
	dI	**0.93 (± 0.29) x 10^{-9}**	**-0.32 (± 0.02)**	**41 ± 3**
	d2-AP/dC	**1.12 (± 0.27) x 10^{-7}**	**-0.18 (± 0.08)**	**23 ± 10**
	d2-AP/dT	**7.89 (± 2.28) x 10^{-9}**	**-0.14 (± 0.04)**	**18 ± 5**
	Iso-dG/Iso-d^{5m}C	**1.35 (± 0.53) x 10^{-8}**	**0.03 (± 0.07)**	**0***
	LA	**6.70 (± 1.93) x 10^{-8}**	**-0.28 (± 0.07)**	**36 ± 9**

Table 3-1 Binding affinities and hydration level of PU.1 and S3 chimera with the modified DNA sequences.

Binding affinities of PU.1 and S3 bound to the various DNAs were measured via fluorescence anisotropy, where the unlabeled DNA of interest was titrated against a Cy3-labeled DNA.

*Binding data was fit utilizing equation (3-1) to provide the dissociation constant. To determine the dependence on the osmotic pressure of the binding affinities, three replicates were used to calculate the averages ± SE. The normo-osmotic affinity was calculated as the equilibrium dissociation constant K_D. The osmotic sensitivity (Γ) of the system is the linear slope of the fitted dissociation constants, as seen in figure 2. The hydration change (Δn_w) is the implied change in hydration of water molecules as determined by equation 3-2 using the osmotic sensitivity (Γ). *Iso-dG/Iso-d^{5m}C resulted in a change in hydration of approximately zero, with the noise there is no discernable difference.*

The first row of the second column illustrates PU.1 bound with inosine substituted DNA (dI) (Figure 3-3). Under normo-osmotic conditions, the PU.1/dI complex is much less intense, and binding affinity decreases ~10 fold. As the osmotic pressure of the system is increased, the PU.1/dI complex is destabilized and results in an uptake of ~72 waters (Table 3-1). Replacing the guanosine with inosine retains the proper interactions between Arg 232 and the carbonyl at position 6 of guanosine. However, inosine loses the minor groove amino group, suggesting that the amino group is critical for retaining high-affinity interactions but not a vital component of the hydration network of PU.1 bound to high-affinity DNA.

PU.1 bound to both d2-AP paired with cytosine (pink) and thymine (green) individually was also examined (Figure 3-3, top row middle panel). PU.1 was first paired with d2-AP/dC to retain the HA complementary sequence while removing the carbonyl from guanosine. Upon performing the binding experiments, it was determined that PU.1 bound to d2-AP/dC performs nearly identically to the LA complex (Table 3-1). However, this raised questions if the binding was poor due to the loss of two hydrogen bonds between d2-AP and dC, the loss of the direct

interaction between Arg 232 and guanosine, or both (Figure 3-3). d2-AP was also paired with dT (green) to test this question, utilizing a dT rather than a dC to retain the two hydrogen bonds (Figure 3-3). PU.1 bound to d2-AP/dT gains ~10 fold binding affinity and is less osmotically sensitive than PU.1 bound to d2-AP/dC, or the PU.1/LA complex. The increase in binding affinity of the d2-AP/dT complex suggests that the wobble pairing of d2-AP/dC is detrimental to the stability of the complex and that by replacing the dC with a dT, it is possible to reestablish some of the binding affinity. However, both the d2-AP/dC and d2-AP/dT complexes prove that the carbonyl on guanosine is essential for maintaining the hydration network iconic to the PU.1/HA complex. The removal of the direct interaction between Arg 232 and guanosine disrupts the hydration network, decreasing hydration to levels comparable to PU.1/LA in d2-AP/dC or lower with d2-AP/dT.

The final modified sequence examined was the Iso-dG/Iso-D^{5m}C complex (Figure 3-3, top row, panel four), in which the Watson and Crick G-C pair is retained with three hydrogen bonds. However, the positions of the carbonyl and the amino groups are swapped. This modification provides a means of exploring the role of the hydrogen bonds between PU.1 and DNA while retaining the proper DNA stability. When paired with PU.1, the loss of the carbonyl on guanosine is essential to maintain high-affinity interactions. Swapping the carbonyl is not enough to compensate for the loss of the direct interaction. Arg 232 on PU.1 cannot compensate or interact with the amino group in the major groove due to a charge mismatch and having double donors. The PU.1 complex with Iso-dG/Iso-d^{5m}C results in LA level interactions, and the loss of hydration is even more significant with only ~17 waters remaining. The hydration of the

Iso-dG/Iso-d^{5m}C complex matches that of the d2-AP/dT complex, proving that the loss of the direct interaction is detrimental to binding affinity and hydration.

3.3.4 Osmotic sensitivity is consistently decreased in the S3 chimera, whereas binding affinity is consistently stronger when bound to high-affinity or modified DNAs.

To better understand the role of hydration in protein-DNA specificity, we utilized a series of chimeras previously examined on high-affinity DNA under increasing osmotic pressure [128]. The chimeras, Loop and S3, were relatively insensitive to osmotic pressure while still retaining PU.1-like binding affinity. These results proved that it is possible to separate the hydration iconic to PU.1 from binding affinity through chimeric constructs [128]. The previous work performed with the Loop and S3 chimeras showed that the Loop retains an intermediate level of hydration compared to PU.1 and S3; for this reason, S3 was chosen as the sole chimera to proceed forward with the modified DNAs. Figure 3-3 shows the fluorescence anisotropy experiments of PU.1 bound to the various DNAs examined vs. S3.

S3 bound to high-affinity DNA (Figure 3-3, bottom row, panel 1) displays the least disruption to binding affinity under normo-osmotic conditions and retains high-affinity binding. The S3/HA complex is significantly less hydrated (~30 waters) compared with the PU.1/HA complex (~90 waters), indicating it is possible to separate the hydration network of PU.1 from high-affinity interactions [128]. In comparison, S3 bound to LA DNA (Figure 3-3 bottom row panel 5) results in an increase in binding affinity while retaining hydration like that of the PU.1/LA complex at ~36 waters [128]. The S3 chimera appears less sensitive to alterations in hydration and sequence, instead binding to HA with decreased hydration while retaining the binding affinity of PU.1/HA and binding LA with stronger binding affinity but similar hydration

levels. As shown previously, Ets-1 residues can reduce the level of hydration of the complex and also decreases the sequence selectivity [128]. This makes S3 an excellent candidate to examine with the modified DNA sequences, providing an avenue to understand better the role of DNA in creating the hydration network of PU.1 and the formation of specific complexes.

S3 was first examined with the dI (blue) sequence (Figure 3-3, bottom row, panel 2). The S3/dI complex retains high-affinity binding while displaying a decrease in hydration to a level that is ~1/2 of what is found in the PU.1/HA complex. Pairing S3 with dI proves that the Ets-1 sequence in the PU.1 scaffolding provides a means of compensation for the loss of the minor groove interaction. The S3 chimera can stabilize the complex enough to retain the HA level of interaction. This stabilization of the S3/dI complex is due to the S3 chimera being less sequence selective than PU.1. Again, the loss of the amino group is not implicated in the hydration network of PU.1, as the loss of the amino group under increasing osmotic pressure is not enough to change the hydration. As expected, the S3/dI complex is less hydrated than the PU.1/dI complex and retains nearly the same hydration as the S3/HA complex (within error, Table 3-1). The corresponding levels of hydration and binding affinity of S3 bound to HA and dI individually provide insight into the location and role of hydration in high-affinity interactions. The lack of change in hydration shows that the hydration network left in the S3/HA complex is located in the major groove between the protein and DNA, rather than split between the major and minor grooves. The retention of the binding affinity in both S3 complexes also displays that the minor groove amino group is not responsible for high-affinity interactions. The Ets-1

chimeras continue to compensate for the loss in specificity enough to retain high-affinity level interactions.

S3 bound to d2-AP with either dC or dT (Figure 3-3, bottom row, panel 3) results in binding better than that of the PU.1/LA complex. However, the S3/d2-AP/dC complex is comparable to the PU.1/d2-AP/dC complex in both binding affinity and level of hydration, which means that the wobble pairing of d2-AP/dC is disruptive enough to overcome the compensation from the Ets-1 sequence. However, the replacement of dC with dT increases binding affinity and decreases hydration when S3 is bound to d2-AP/dT. The S3/d2-AP/dT complex displays a stronger binding affinity than PU.1/d2-AP/dT while retaining comparable hydration levels. The increased binding affinity is again explained by using Ets-1 residues in the place of PU.1 residues, providing a means to compensate for the loss of the direct interaction between Arg 232 and the carbonyl of guanosine through more favorable sidechain interactions. The removal of the direct interaction also disrupts the hydration network of the S3/HA complex, dropping hydration nearly in half from ~31 waters to ~18.

The final modified guanosine examined by S3 is the Iso-dG/Iso-d^{5m}C, and as expected, the binding affinity decreased but just to the level of the S3/d2-AP/dT, not entirely to the S3/LA. This suggests that the S3 chimera responds to the carbonyl/amino group swap in the same manner as removing the carbonyl. The decrease in binding affinity was not as drastic as the decrease displayed by PU.1. The chimera continues to compensate for less favorable interactions, even with extreme modifications like a carbonyl/amino group swap in the core consensus sequence.

The S3/Iso-dG/Iso-d^{5m}C is notable in hydration- the complex is the least hydrated complex of all the combinations examined thus far, with hydration at nearly 0. This suggests that

the pairing of S3 with Iso-dG/Iso-d^{5m}C almost completely knocks down the hydration network of PU.1. The decrease in hydration is expected, as the S3/HA complex decreases the hydration level by 2/3, and the loss of the carbonyl with S3 drops the hydration level by another ½. The almost complete removal of hydration by the combination suggests that S3 can stabilize the complex along the 5' flanking region of DNA without directly interacting with the modified DNA while losing the remaining waters. This suggests a possible allosteric mechanism in which S3 stabilizes H3 on the modified DNA, providing a means for increased binding affinity with a nearly complete loss in hydration.

3.3.5 Methylation of the complementary cytosine decreases binding affinity while retaining hydration, suggesting that swapping the positions of the carbonyl and amino groups successfully knocks down the hydration network of PU.1

When designing the Iso-dG sequence, the desire was to pair it with an Iso-dC to retain the same chemical makeup as the HA pairing while swapping the positions of the carbonyl and amino groups. However, it was not possible to acquire an Iso-dC. Instead, an Iso-d^{5m}C was utilized in which position 5 on the cytosine was methylated in addition to the carbonyl and the amino group swap. To determine that the trends displayed when PU.1 was bound to Iso-dG/Iso-d^{5m}C was in fact due to the carbonyl/amino group swap and not due to the methylation, we performed a control experiment in which we paired guanosine with a 5-methylated cytosine (Figure 3-4, dG/d^{5m}C, purple). The fluorescence anisotropy experiments with the methylated cytosine show the binding affinity is higher than that in Iso-dG/Iso-d^{5m}C, and there is an increase in hydration (Figure 3-4, panel 1). This shows that although methylation of cytosine is disruptive, it is not as disruptive as when the carbonyl and amino groups have swapped. Through

the carbonyl/amino group swap, the hydration network of PU.1 is entirely knocked down. This data proves that retaining the proper Watson and Crick interaction between dG and dC is not enough to maintain the hydration network or the binding affinity of the HA complex. This also shows that the hydration network requires a hydrogen bond acceptor at the position of the second guanosine in the consensus sequence; by swapping with a hydrogen bond donor, it is not possible to recapture the hydration network. The PU.1/dG/d^{5m}C complex was found to have a hydration

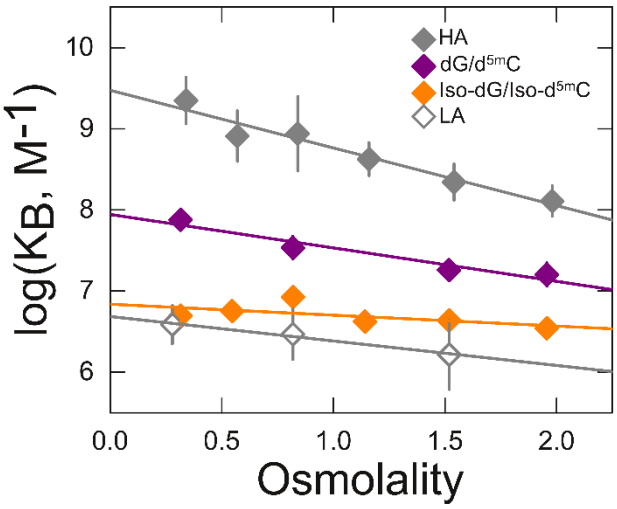


Figure 3-4 Methylation of the complimentary strand of DNA results in increased hydration and binding affinity as compared with Iso-dG/Iso-d^{5m}C.

A comparison of PU.1 bound to Iso-dG/Iso-d^{5m}C (orange) and dG/d^{5m}C (purple). The addition of a methylation on C5 of cytosine causes a decrease in binding affinity with the partial retention of the hydration network of PU.1. However, when paired with an Iso-dG/Iso-d^{5m}C, binding affinity is decreased to LA levels, and the hydration network is completely knocked down. The methylation of cytosine results in a normo-osmotic pressure binding affinity that is about halfway between the high (solid grey) and low-affinity (open grey) complexes, with a level of hydration that is decreased by 1/3 when compared with the HA complex. Meaning that the methylation is disruptive to the level of hydration and binding affinity of the HA complex but is not as detrimental as the swapping of the positions of the amino and carbonyl groups.

level 1/3 less than the HA complex, and the level of binding affinity under normo-osmotic pressure was about halfway between the HA and LA complexes (Table 3-2).

		normo-osmotic affinity (K_d, M)	***osmotic sensitivity (Γ, osm^{-1})***	***Hydration change (Δn_w)***
PU.1	**HA**	**0.34 (± 0.06) x 10^{-9}**	**-0.70 (± 0.05)**	**90 ± 6**
	Iso-dG/Iso-d^{5m}C	**2.03 (± 0.32) x 10^{-7}**	**-0.13 (± 0.08)**	**17 ± 10**
	dG/d^{5m}C	**1.32 (± 0.08) x 10^{-8}**	**-0.41 (± 0.07)**	**52 ± 9**
	LA	**2.59 (± 1.41) x 10^{-7}**	**-0.30 (± 0.04)**	**38 ± 5**

Table 3-2 A comparison of the binding affinities and level of hydration of PU.1 bound to HA, Iso-dG/Iso-d^{5m}C, dG/d^{5m}C, and LA.

Competition fluorescence anisotropy experiments determine binding affinities under normo-osmotic conditions without the addition of any osmolytes to determine the binding constant, as explained in Table 3-1. Osmotic sensitivity (Γ) comes from the linear slope of the change in binding affinity vs. the difference in osmotic pressure from two to three replicates used to calculate the average and SE. The hydration change (Δn_w) is calculated using equation 3-2 and calculates the change in hydration of water molecules using osmotic sensitivity (Γ).

3.3.6 PU. 1's sequence selectivity is driven by exothermic reactions and increased levels of hydration.

The fluorescence anisotropy experiments of PU.1 and S3 bound to HA, LA, and modified DNA sequences have provided increased insight into how the modifications alter sequence selectivity, binding affinity, and hydration to provide a more substantial view of how hydration drives sequence selectivity and binding affinity. However, the binding data does not provide any information on the thermodynamics of protein/DNA interactions. To better understand the thermodynamics of the protein/DNA complex, we have chosen to examine a select number of

combinations by Isothermal Calorimetry Titration (ITC). We will be focusing on the most structurally stable combinations and the most exciting changes in hydration.

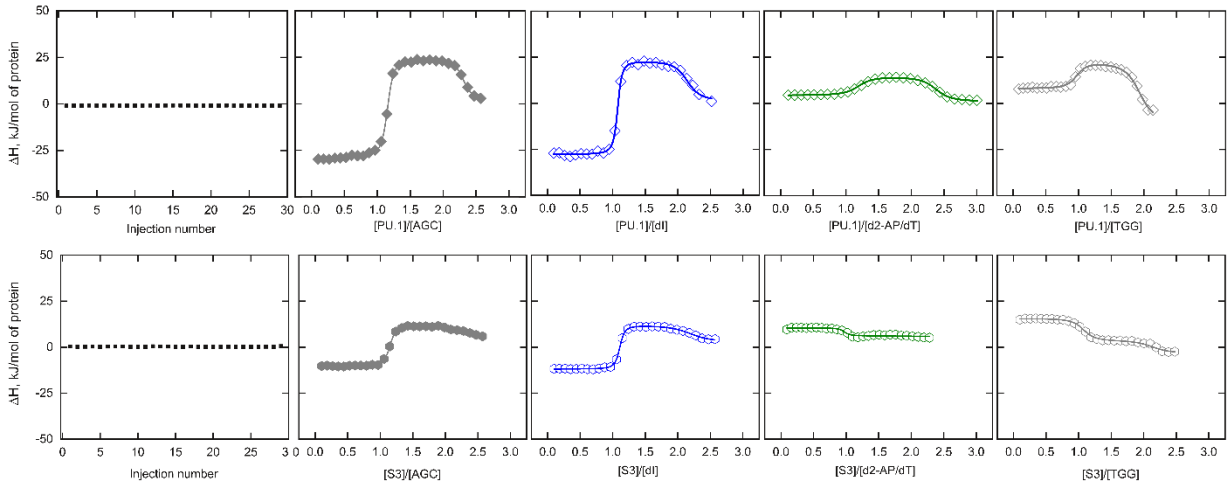


Figure 3-5: Measured changes in enthalpy of PU.1 and S3 bound to high-affinity, low-affinity, and modified DNA provide insight into the thermodynamic effects of modified deoxyguanosine bases.

Top row: Panel (1) PU.1 was injected into a well containing buffer alone as a control to demonstrate that any changes in enthalpy are due to interactions between the protein and DNA and not the buffer. Panel (2) PU.1 binds HA by first forming a 1:1 complex of protein on DNA through an exothermic reaction, followed by the formation of a 2:1 complex where two PU.1 molecules are bound to one DNA sequence. The formation of the 2:1 complex results from an endothermic reaction. Panel (3) PU.1 bound to dI displays a remarkably similar thermodynamic pattern to PU.1/HA complex. Panel (4) Upon binding d2-AP/dT, there is a switch in the change in enthalpy. Forming the 1:1 complex of PU.1/d2-AP/dT becomes endothermic, and the formation of the 2:1 complex results in even more loss of heat. Panel (5) is PU.1 binding LA, PU.1 forms both the 1:1 and 2:1 complexes endothermically. Row 2: Panel (1) S3 was injected into buffer alone and displayed no change in enthalpy. In panels (2 and 3), S3 binds HA and dI DNAs similarly, with an exothermic change in enthalpy upon forming the 1:1 complex and an

endothermic change in energy to form 2:1 complexes. S3 bound to both HA and dI follow the same binding patterns as PU.1 with the same DNAs. In panels (4 and 5), S3 also forms the 1:1 complex with d2-AP/dT and LA individually by an endothermic reaction. However, for the 2:1 complex, instead of an increase in an endothermic reaction, there is a decrease, thus diverging from the pattern displayed by PU.1.

The combinations we chose to examine more closely by ITC include PU.1 bound with HA, dI, d2-AP/dT, and LA individually, as well as S3 bound with HA, dI, d2-AP/dT, and LA also separately (Figure 3-5). Upon performing these titrations, exciting patterns started to emerge. One striking pattern was that the more hydrated the complex, the more significant the change in enthalpy (ΔH1) of the first binding event. The PU.1/HA complex is the most hydrated at ~90 waters (Table 3-1) and has the most significant ΔH1 at -31 kJ/mol at 25°C (Figure 3-5 and 3-6, Table 3-3), along with the most significant change in Gibbs free energy (ΔG) at -54 kJ/mol at 25°C (Figure 3-5 and 3-6, Table 3-3).

$$\Delta G = RT \ln K_d$$ Equation 3-3

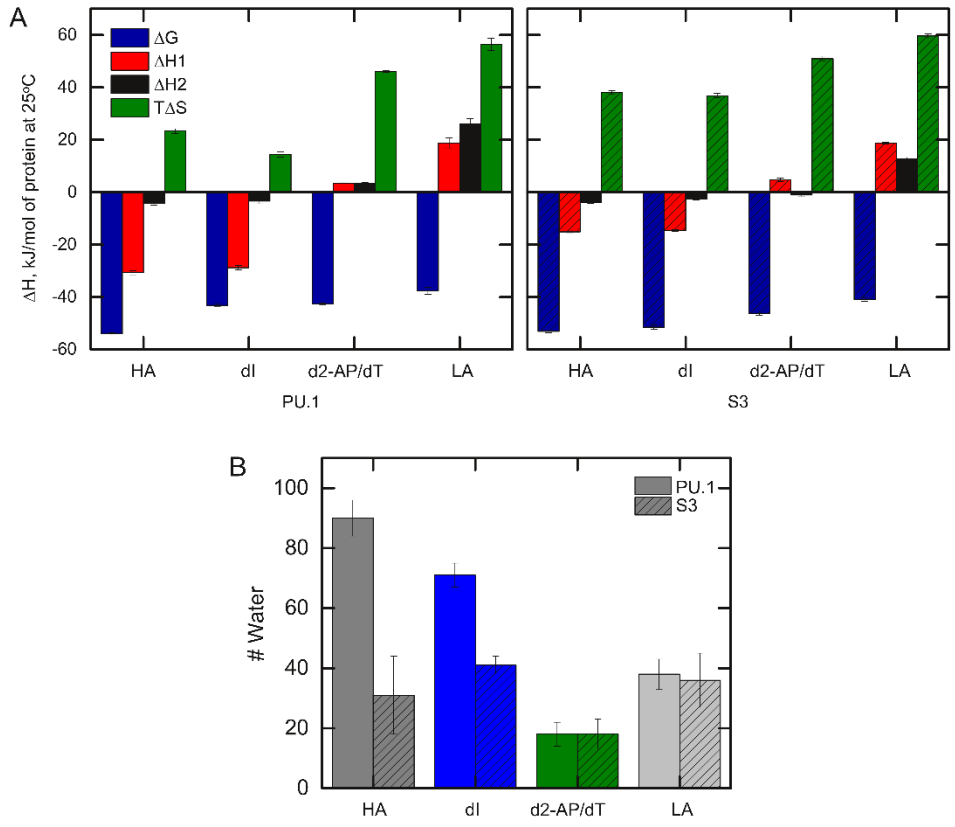


Figure 3-6: Sequence selectivity drives thermodynamic properties of protein/DNA interactions.

The change in enthalpy (ΔH) is directly measured through ITC. The measurement of ΔH provides a means of calculating the change in Gibbs free energy (ΔG) and the change in entropy (TΔS) of the system. This information provides greater insight into the role hydration, and sequence selectivity plays in protein/DNA interactions. A) Shows the calculated and measured thermodynamic properties of PU.1 bound to HA, dI, d2-AP/dT, and LA DNAs (solid colors) as well as S3 bound to the same DNA sequences (stripes). B) A comparison of the hydration of each of the PU.1(solid colors) and S3 (striped) complexes examined by ITC.

The more negative the ΔG is, the more spontaneous/favorable the reaction will be; as expected, PU.1/HA is the most favorable reaction. This reiterates the observation from the

fluorescence binding experiments above. ΔG also provides a means of determining whether a protein binds a sequence specifically or non-specifically. To determine the differences in sequence specificity, the change in the change of Gibbs free energy (ΔΔG) was calculated (equation 3-4):

$$\Delta\Delta G = RT \ln K_d \left(\frac{A}{B} \right) \qquad \text{Equation 3-4}$$

Using this ΔΔG provides valuable insight into a protein's specificity for varying DNA sequences, a protein that binds strongly to a variety of sequences will have a small ΔΔG when comparing those sequences but will be found not to be sequence specific. A protein with sequence specificity will bind strongly to a particular sequence rather than binding well to various sequences.

kJ/mol at 25°C		***ΔG***	***ΔH1***	***ΔH2***	***TΔS***
PU.1	**HA**	-54.0 ± 0.4	-30.7 ± 0.8	-4.1 ± 0.9	23.4 ± 0.9
	dI	-43.1 ± 0.4	-28.8 ± 0.9	-3.4 ± 1.0	14.4 ± 1.0
	d2-AP/dT	-42.6 ± 0.2	3.4 ± 0.2	8.5 ± 0.2	46.0 ± 0.3
	LA	-37.6 ± 1.4	18.8 ± 1.9	26.0 ± 2.0	56.4 ± 2.3
S3	**HA**	-53.1 ± 0.6	-15.1 ± 0.3	-3.9 ± 0.5	38.0 ± 0.7
	dI	-51.5 ± 0.8	-14.6 ± 0.3	-2.7 ± 0.3	36.9 ± 0.8
	d2-AP/dT	-46.2 ± 0.7	4.7 ± 0.6	2.7 ± 1.2	50.9 ± 1.0
	LA	-40.9 ± 0.7	18.7 ± 0.4	12.8 ± 0.4	59.7 ± 0.8

Table 3-3 Thermodynamic parameters of PU.1 and S3 bound to HA, LA, and modified DNAs determined by ITC

The change in Gibbs free energy (ΔG) is calculated through the equilibrium dissociation constant using equation 3-3. It provides insight into the favorability of the reaction, as a negative ΔG results in a spontaneous reaction. ITC provides a means of directly measuring the change in enthalpy (ΔH), or the heat of reaction, of a protein interacting with DNA. In these experiments, the well contains DNA, and protein is titrated into the well; as the protein binds the DNA, heat is

either released or taken up from the surrounding environment, measured by the instrument. The titration beings with a surplus of DNA and is continued until a molar ratio of at least 2 molecules of protein are present to 1 mole of DNA, which allows for the measurement of ΔH *for the forming of the 1:1 complex and the 2:1 complex. Once* ΔG *has been calculated and* ΔH *has been measured, it is then possible to measure the change in entropy at a specific temperature (*$T\Delta S$*) via equation 3-4. A fuller picture of the thermodynamic profile is developed for each of these systems through the measurement of* ΔH *and the calculation of* ΔG *and* $T\Delta S$*.*

PU.1 bound to modified or LA DNA resulted in decreased binding affinity, as expected by PU.1's highly selective targeting for specific sequences (Figure 3-7) [73]. Of the sequences examined when bound to PU.1, dI was the most thermodynamically similar to PU.1/HA. The PU.1/dI complex was found to form the 1:1 complex with a ΔG slightly lower than the PU.1/HA complex (-43.1 kJ/mol at 25°C vs. -54.0 kJ/mol at 25°C) but retained similar changes in enthalpy for both the 1:1 and 2:1 complex (Table 3-3). The binding experiments show that PU.1 binds dI with lower affinity and lower levels of hydration. The lower affinity would affect the spontaneity of the complex's formation, as this is a less favorable interaction and thus requires more energy to form a stable complex. It is surprising that even with the decreased levels of hydration (PU.1/HA ~90 waters and PU.1/dI ~72 waters), the changes in enthalpy of the 1:1 complexes (PU.1/HA ~ -31 kJ/mol at 25°C and PU.1/dI ~ -29 kJ/mol at 25°C) are still similar between the two sequences. Between the PU.1/HA and PU.1/dI complexes, the ΔG becomes less negative by ~10 kJ/mol at 25°C, and the $T\Delta S$ also becomes less positive by ~10 kJ/mol at 25°C.

$$\Delta G = \Delta H - T\Delta S$$ Equation 3-5

The changes in ΔG and TΔS and similar ΔH indicate that the waters lost in the PU.1/dI complex are not enthalpically important but instead alter the system's entropy (TΔS). The decrease in hydration decreases the system's entropy, which is not expected as a release of water is expected

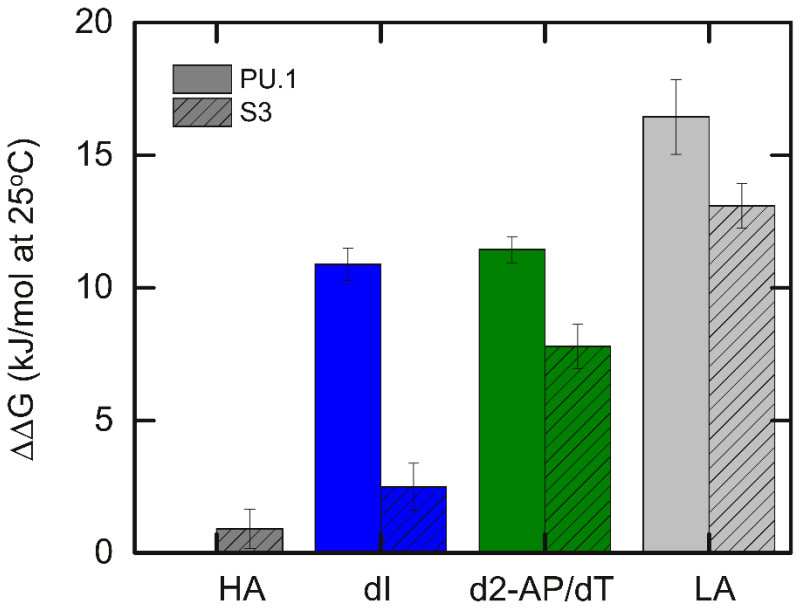


Figure 3-7 S3 is less sequence specific than PU.1 The change in the change in Gibbs free energy (ΔΔG) was calculated using the binding affinity of PU.1 bound to HA as a comparison. The PU.1/HA complex was selected as the reference due to the complex forming with the strongest binding affinity. This provides a means of comparing the sequence selectivity between PU.1 and S3 bound to the various DNA sequences examined. S3 consistently has a smaller ΔΔG when compared to PU.1, forming the same complex. The differences in ΔΔG explains that the addition of Ets-1 residues in the PU.1 scaffolding provides a means of increasing the proteins' ability to bind effectively to a variety of sequences but is less sequence selective. In contrast, PU.1 has a high sequence selectivity, as displayed by the rapidly growing ΔΔG.

to raise the system's entropy rather than reduce it [28]. This decrease in entropy paired with a reduction in hydration raises questions of if a structural change occurs to release waters, but also decrease the level of entropy of the system?

The S3 chimera is consistently less sequence selective than PU.1 (Figure 3-7). However, S3 bound to HA and dI sequences show similar thermodynamic patterns to PU.1 with identical sequences. S3 shows the greatest ΔG when bound to HA DNA, at -53 kJ/mol at 25°C, which is the same as the ΔG of the PU.1/HA complex, within error (Figures 3-5 and 3-5, Table 3-3). The ΔH1 is half as negative in the S3/HA complex than in the PU.1/HA complex (Figures 3-5 and 3-6, Table 3-3). This difference in ΔH1 can be attributed to the difference in the number of waters found in the complex. The S3/HA complex retains ~31 waters to the ~90 waters located in the PU.1/HA complex (Table 3-1, Figure 3-6). These results agree with the trend displayed by PU.1, where the PU.1/HA complex had a more negative ΔH1 than the less hydrated PU.1/dI complex. S3 starts to diverge thermodynamically in the S3/dI complex, which has ~10 more waters than the S3/HA complex. The S3/dI complex has a similar ΔG to that of both PU.1 and S3 bound to HA, at -51.5 kJ/mol at 25°C, and although being slightly more hydrated than the S3/HA complex, the ΔH1 of S3/dI is nearly identical at -14.6 kJ/mol at 25°C (Figures 3-5 and 3-6, Table 3-1 and 3-3). TΔS of the S3/dI complex is also similar to that of the S3/HA complex at 36.9 kJ/mol at 25°C. While slightly more hydrated, The S3/dI complex thermodynamically is very similar to the S3/HA complex (Table 3-3), suggesting that the sequence selectivity of S3 drives the thermodynamic properties. Since S3 is less sequence selective, the S3 complexes with HA and dI have less pronounced differences in thermodynamics (Figure 3-7).

However, when comparing the PU.1/dI complex with the S3/dI complex, we find that the ΔG of the S3/dI complex is more negative than the PU.1/dI complex by 8 kJ/mol at 25°C (Table

3-3). The ΔH1 is more negative in the PU.1/dI complex, whereas the TΔS is more positive for the S3/dI complex. This is expected as the PU.1/dI complex is more hydrated by ~30 waters. Hydration plays a prominent role thermodynamically, as seen in the PU.1/HA and PU.1/dI complexes, and the trend continues in the S3/dI complex. The decreased hydration of the S3/dI complex increases the system's entropy by not having to expend energy to retain the extra ~30 waters.

3.3.7 Low-affinity protein/DNA interactions are driven by larger positive changes in entropy and positive changes in enthalpy regardless of hydration.

As PU.1 and S3 bind to DNA sequences with lower binding affinities and less specificity, ΔG becomes less negative and less favorable, ΔH1 becomes less negative, and TΔS becomes more positive. As sequences with lower binding affinities tend to have less interfacial hydration and higher entropy requiring less energy to restrain localized waters, which are instead released into bulk solution. As the systems move from high-affinity to low-affinity interactions, ΔH1 changes signs to indicate high-affinity interactions are exothermic (-ΔH) or release heat upon complex formation. In low-affinity interactions, the thermodynamic properties change from exothermic to endothermic (+ΔH) or an uptake of heat from the surrounding environment. This change in sign on ΔH1 signals the thermodynamic difference between high-affinity and low-affinity interactions.

This change in ΔH1 results from either the loss of the hydration network in PU.1 or from an unfavorable DNA sequence. The PU.1/HA complex forms a hydration network, while the PU.1/LA complex is characterized by decreased hydration. The release of water increases the entropy and increases the amount of heat taken up by the system. The PU.1/d2-AP/dT complex is an intermediate between the HA and LA complexes in binding affinity. Thermodynamically,

PU.1/d2-AP/dT is also between the PU.1/HA and PU.1/LA complexes. Since the binding affinity of PU.1/d2-AP/dT is closer to the LA complex than HA, we expect to see the thermodynamics to be more like low-affinity. This is precisely what was found- ΔG of PU.1/d2-AP/dT at 25°C is comparable to the LA complex (within 5 kJ/mol) and ~11 kJ/mol less negative than the HA complex. In the PU.1/d2-AP/dT complex, ΔH1 (3.4 kJ/mol) is closer to the LA complex (18.8kJ/mol) than to the HA complex (-30.7 kJ/mol). These changes in the ΔG and ΔH1 values correspond to rank order decrease in binding affinities between the three complexes- relative to the optimal sequence. Modifying guanosine to d2-AP results in decreased binding affinity and hydration, resulting in the complex becoming closer to low-affinity in terms of thermodynamics.

However, the LA complex is more hydrated than d2-AP/dT, suggesting LA should be more thermodynamically favorable than d2-AP/dT; this is not the case. The defining characteristic of low-affinity sequences is in the flanking regions of the cognate site. The LA sequence retains the proper 5' flanking region and the core consensus but varies at the 3' flanking region. The HA 3' flanking sequence consists of a 5'-GTG-3', whereas the LA complex has a 5'-TGG-3' instead. The swap of the thymine and guanosine in the flanking region may play a more decisive role in the differences in the thermodynamic pattern than the hydration. Previous work on high-affinity DNA sequences has shown that changing the DNA bases at flanking positions can significantly differ in thermodynamic parameters like ΔH and TΔS [84]. Even though the PU.1/LA complex has more hydrating waters than PU.1/d2-AP/dT, the change in bases is more thermodynamically disruptive than the difference in waters.

S3 bound to d2-AP/dT and LA DNA follows the same pattern as PU.1 bound to the same sequences. The higher binding affinities correspond to more negative ΔG, ΔH1, and the less positive the TΔS, but S3/ d2-AP/dT or S3/LA undergoes a flip with ΔG becoming less negative.

In contrast, both ΔH1 and TΔS become more positive. The S3/d2-AP/dT complex is less hydrated than its LA counterpart but remains thermodynamically between S3/HA and S3/LA complexes. The S3/LA complex is more hydrated than S3/d2-AP/dT. The trend continues to be similar- ΔG is less negative, ΔH1 and TΔS are more positive, which may be attributed to the difference in the 3' flanking sequence rather than hydration.

However, we see a notable trend when comparing the thermodynamic properties between the PU.1 and S3 complexes. As expected, ΔG is most negative for PU.1/HA, but ΔG of S3/d2-AP/dT and S3/LA are more negative than the corresponding complex with PU.1. The only difference between the complexes, other than protein, is the level of hydration, with S3 being consistently less hydrated than PU.1. We know that S3 is less discriminating against DNA from the binding data, ΔΔG, and can bind to a broader variety of modified DNAs than PU.1. This decrease in specificity but increase in affinity appears to correspond with more favorable ΔG, smaller ΔΔG and suggests interactions that compensate for hydration.

3.3.8 Contributions from hydration and DNA dynamics can be separated through the use of nucleotide derivatives

The guanosine derivatives chosen provide a means of altering DNA in precise mechanisms. The Iso-dG derivative retains DNA stability through the retention of the three hydrogen bonds between Iso-dG and Iso-^{5m}dC. In contrast, dI results in the removal of the minor groove amino group, which provides a mechanism for destabilizing the DNA away from the protein-DNA interface. The removal of the amino group in the minor groove removed one of the three hydrogen bonds between guanosine and cytosine and was previously found to open the minor groove slightly, suggesting that with the opening of the minor groove, DNA becomes more dynamic in that position [132]. The dI substitution is interesting as it increases the DNA

dynamics but does not affect the level of hydration of the complex, as displayed by both PU.1 and S3 (Figure 3-3 and Table 3-1). Whereas, in comparison, d2-AP may also increase DNA dynamics due to the loss of the carbonyl and the wobble pairing of d2-AP/dC or the two hydrogen bonds in d2-AP/dT but does not retain the hydration network of the PU.1/HA complex.

Focusing on the dI complexes with PU.1 and S3 provides valuable insight into the role of hydration and DNA dynamics. PU.1 was found to retain higher levels of hydration when bound to dI DNA but suffers from decreased binding affinity under normo-osmotic conditions. In comparison, S3 forms a less hydrated complex with dI while still retaining high-affinity interactions under normo-osmotic conditions. These results suggest that the hydration network of PU.1 may have evolved as a means of restraining DNA dynamics when forming a specific complex. S3 is less hydrated and is thus able to bind with stronger binding affinities to more dynamic DNAs like dI. Whereas PU.1 retains the hydration network and is less able to compensate for the increased DNA dynamics caused by removing the amino group in the minor groove. The hydration network of PU.1 appears to compensate to an intermediate level for the increased DNA dynamics, as the level of binding affinity for PU.1/dI is in between the HA and LA complexes. S3 is less sensitive overall to changes in DNA dynamics, as it retains the ability to bind with stronger affinities to all the DNA sequences examined when compared with PU1 (Figure 3-3 and Table 3-1). Whereas PU.1, when bound with d2-AP/dC or d2-AP/dT, results in drastic losses in both hydration and binding affinity. The use of the nucleotide derivatives has provided an avenue for separating hydration from DNA dynamics, as demonstrated by the S3 chimera paired with dI DNA.

3.3.9 PU.1 and S3 form 2:1 complexes on modified DNA specifically

ITC measurements provide valuable data on whether the protein of interest successfully binds to the DNA and whether it binds as a 1:1, 2:1, or a higher-order complex. ITC also provides valuable information on whether the protein is binding specifically or not with DNA. In Figure 3-5, a series of ITC plots show PU.1 or S3 titrated with different DNA sequences. The experiment starts with a set concentration of DNA and titrates in an increasing concentration of either PU.1 or S3, which binds the DNA. Performing the experiment in this manner provides a means to determine the binding mode of the protein for the DNA. The slow addition of the protein to the DNA allows the protein to form the 1:1 complex first, followed by the 2:1 complex. In all the combinations examined, the protein and DNA formed a specific 1:1 complex. All but one set (S3/d2-AP/dT) formed a specific 2:1 complex upon increasing the protein's concentration.

The specificity of the complex is determined by whether there is a change in the heat of the reaction at the molar ratio expected. For the first binding event in all cases, ΔH starts either below or above 0 kJ/mol at 25°C, and increasing the protein concentration in the well to a molar ratio of 2 will either undergo another binding event (2:1) or will return to 0 kJ/mol at 25°C. In the formation of nonspecific complexes, the addition of the protein would continue to increase or decrease the ΔH, failing to return to 0 kJ/mol at 25°C. All of the complexes examined with PU.1 form clear 1:1 and 2:1 complexes and returned to zero with the addition of more protein after the second binding event. This supports the proposal that PU.1 can form both the 1:1 and 2:1 complexes specifically with the modified DNA sequences. When examining S3 on the modified DNA sequences, we find three of the four sequences form consistent 1:1 and 2:1 specific complexes: HA, dI, and LA. However, when bound to d2-AP/dT, there is a clear 1:1 binding

event and a decrease in ΔH following. After the point of the 2:1 binding event, only a slight change occurs, which could be the transition from the 2:1 complex back to zero or a nonspecific binding event. More data is required for a clear answer regarding if S3 binds d2-AP/dT in a specific 2:1 complex.

3.4 Discussion

Both PU.1 and the S3 chimera retain the ability to bind with the HA, LA, and modified sequences to varying degrees. PU.1 is consistently destabilized when bound to the modified or LA DNAs but retains much of the water network in the PU.1/dI complex. When bound to the modified DNAs individually, PU.1 binds like PU.1/LA, with decreased hydration levels and decreased binding affinities. The S3 chimera binds each of the sequences examined with increased binding affinity relative to the corresponding PU.1 complexes, but with reduced hydration. S3 follows the same pattern as displayed by PU.1. When binding the modified DNA sequences d2-AP/dC, d2-AP/dT, and Iso-dG/Iso-d^{5m}C, they form a low-affinity type complex with decreased hydration levels and decreased binding affinity. The S3 chimera was less sequence-specific and therefore bound each of the DNAs with stronger binding affinities. The loss of the hydration network, that is crucial to PU. 1's sequence selectivity, in the S3 complex likewise leads to a decrease in sequence sensitivity [73, 128].

3.4.1 Direct readout is essential for high-affinity interactions, whereas indirect readout is vital for increased specificity of sequence selection.

The introduction of the modified nucleobases dI, d2-AP, and Iso-dG/Iso-d^{5m}C provide a means of probing the role of the DNA in forming high-affinity-specific complexes. By pairing these sequences with PU.1 and S3, it was possible to determine the effect of changing the base readout on binding specificity. These modified bases provided a means of exploring direct and

indirect readout individually and together on protein/DNA interactions. PU.1 is the more sequence selective of the two proteins examined; when paired with dI to modify indirect readout, fluorescence anisotropy indicated the loss of the minor groove interaction decreased binding affinity without significant impact to the hydration network. S3 paired with dI results in high-affinity interactions and similar hydration levels as the S3/HA complex. From these results, we conclude that the more sequence-specific the protein is, the more indirect readout plays a role in sequence recognition. However, decreasing the rate of sequence selectivity results in a decrease in the effect of altering indirect readout.

Removing direct readout mechanisms, such as with the d2-AP constructs, in which the direct hydrogen bond between Arg 232 and guanosine is removed, is detrimental to both binding affinity and the hydration network of PU.1. The hydration network of PU.1 was determined to be located in the major groove of the DNA in the protein-DNA interface and is critical to retaining high-affinity level interactions. Using the d2-AP constructs, we found that removing the direct interaction between the DNA and PU.1 destroys the hydration network, dropping the hydration level from ~90 waters down to ~31 waters with d2-AP/dC and ~18 waters with d2-AP/dT. Along with the disruption of the hydration network, the loss of the direct interaction also significantly inhibits binding, dropping binding affinity to a level similar to low-affinity. S3 bound to the same sequences shows similar levels of hydration with a slight increase in binding affinity. These results show that the evolutionarily conserved interaction between Arg 232 and guanosine is essential for high-affinity interactions and the retention of the hydration network. The loss of the direct interaction also clarifies again that more sequence-specific proteins are less tolerant to modified bases, as PU.1 binds the modified sequences containing d2-AP slightly weaker than S3.

Swapping the hydrogen bond acceptor and donor positions is insufficient to retain either high-affinity interactions and hydration. The Iso-dG/Iso-d^{5m}C construct retains proper Watson and Crick base pairing but in an isoform setup, allowing the DNA to remain stable but swapping the position of the carbonyl and amino groups in the major and minor groove of duplex DNA. The use of Iso-dG/Iso-d^{5m}C shows that DNA stability is less of a factor for retaining high-affinity binding. Instead, the location and ability to hydrogen bond properly are required for proper high-affinity interactions. Neither PU.1 nor S3 maintain the ability to create the hydration network needed, proving that the carbonyl is required at carbon 6 on guanosine to form the hydration network. DNA readout is essential for forming high-affinity protein/DNA complexes; disruptions to the readout, especially direct readout, are highly disruptive to the formation of the complex.

3.4.2 Differing thermodynamic properties drive sequence specificity

High-affinity protein/DNA complexes are driven by exothermic binding reactions upon forming the 1:1 complex, whereas endothermic reactions characterize the formation of low-affinity complexes. This change in the enthalpy sign provides an apparent division between high and low-affinity complex formation. PU.1, being the more sequence-specific protein examined here, displays more considerable enthalpy changes for high-affinity complexes than the less sequence-specific construct S3. PU.1 is characterized as more exothermic, whereas S3 is more endothermic than PU.1 bound to low-affinity sequences. This suggests increased sequence specificity undergoes an increased release of heat upon binding. In contrast, the less sequence-specific protein will have a larger uptake of heat upon binding.

With the decrease in specificity, there is also a change in the Gibbs free energy of the system. The PU.1/HA complex displays the most negative and, therefore, most favorable reaction, with the S3/HA complex following close behind. As PU.1 loses binding affinity when bound to less favorable DNA sequences, the negativity of ΔG decreases and the ΔG of S3 becomes more favorable than its PU.1 counterparts. This change in ΔG shows that S3 retains the ability to bind more sequences with higher binding affinities and is thus less sequence selective than PU.1. The difference in thermodynamic properties provides a valuable understanding of the differences in sequence selectivity between the proteins and DNA sequences examined here.

3.4.3 PU.1 hydration has evolved as a means of restraining DNA dynamics

Deoxyguanosine derivatives have provided a means of separating the role of hydration from DNA dynamics. Pairing PU.1 and S3 with dI individually has provided insight into the role of hydration in response to increased DNA dynamics. The introduction of the dI into the consensus sequence results in the loss of a hydrogen bond between inosine and cytosine due to the removal of the amino group on guanosine, this, in turn, acts to increase the dynamics of the DNA at that position due to the lost interaction and the decreased lifetime of the bond [132]. Pairing PU.1 with dI resulted in decreased binding affinity while retaining the hydration network. Whereas S3 paired with dI resulted in high-affinity level interaction while not affecting the hydration network. The pairing of S3/dI displays that it is possible to retain high-affinity level interactions on DNA with increased dynamics with the decreased hydration network. In contrast, pairing PU.1/dI with the retention of the hydration network results in decreased binding affinity. Which provides a picture of the role of hydration with DNA dynamics, in which the hydration network allows the protein to interact with DNA with increasing dynamics up until a

certain point, in which the dynamics becomes too large for the hydration to restrain. The PU.1/dI complex retains a binding affinity intermediate to the HA and LA complexes, suggesting that the hydration network does restrain the DNA dynamics to an intermediate level. The less hydrated S3 is less sensitive to the increased DNA dynamics, as displayed by the high-affinity level of interaction. S3 is consistently less sensitive to increased DNA dynamics when compared with PU1, as seen by the increased binding affinity when bound to d2-AP/dC and d2-AP/dT.

3.5 Conclusion

Both PU.1 and S3 bind the HA DNA with comparable binding affinities under normo-osmotic pressure but with drastically different patterns of interfacial hydration when examined under increased osmotic stress. Which drove the question of whether the DNA is playing a significant role in determining the level of binding affinity and hydration of the protein/DNA complex. Through analysis with modified deoxyguanosine derivatives, it was determined that the more sequence selective a protein is for DNA, the more direct and indirect readout affects binding. The difference in sequence selectivity is also displayed through differing levels of hydration and differing thermodynamic patterns. Modified deoxyguanosine derivatives also provided a means to separate the role of hydration from DNA dynamics, allowing for a greater understanding of how DNA dynamics affects the formation of protein/DNA complexes. By using the structurally conserved ETS domain chimeric constructs, it was possible to probe the role of DNA in sequence selectivity without modifying the recognition helix. Chimeric constructs provide a valuable means of assessing the functional properties of protein/DNA interactions.

4 CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Increased levels of hydration drive sequence selectivity

Studying the ETS family of transcription factors has provided a means to examine the role of hydration in protein/DNA interactions due to the structural conservation of the DNA binding domain throughout the family. Ets-1 and PU.1 were chosen due to the varying levels of hydration found in the crystal structures, the conserved DNA binding domain, and their highly divergent primary sequences. PU.1 was previously found to retain a more significant number of waters within the DNA binding interface, and this water was found to be crucial for high-affinity interactions with DNA [128]. By increasing the osmotic pressure of the system, the complex is destabilized, demonstrating that the hydration network of PU.1 is essential for high-affinity interactions [128]. Analyzing the complex under osmotic stress provides a means of determining the number of water molecules taken up or released when forming the protein/DNA complex. This project aimed to better comprehend hydration in the context of protein/DNA interactions and why evolutionary pressures caused PU.1 to develop this extensive hydration network in high-affinity protein/DNA complexes.

Through this research, we have determined that hydration in PU.1 was evolutionarily selected as a means of sequence selectivity rather than an essential requirement for high-affinity interactions. By replacing individual structural elements in PU.1 with the corresponding structure in Ets-1, we proved that it is possible to retain high-affinity interactions as long as the structural framework is maintained [128]. With the ability to retain high-affinity interactions, it was also determined that through swapping secondary structures, it is possible to cause a significant decrease in both hydration and osmotic sensitivity in the PU.1 chimeras [128]. The decreased levels of hydration found in the chimeric constructs binding a range of DNA sequences was

coupled with a decrease in sequence selectivity. The S3 chimera is a prime example- replacing PU.1's S3 with the corresponding sequence of Ets-1 caused the S3 chimera to bind high-affinity DNA with the same binding affinity as PU.1, but also increased binding to low-affinity DNA. This desensitization demonstrates that S3 is less sensitive to DNA sequences than PU.1 and is thus less sequence selective.

The decrease in sequence specificity of S3 vs. PU.1 was displayed again when testing binding to modified DNA sequences containing inosine (dI), 2-Aminopurine (d2-AP), and iso-deoxyguanosine (Iso-dG). S3 retained a high-affinity level of interaction with the dI DNA, retaining and demonstrated consistently stronger binding affinities to the other modified DNAs than displayed by PU.1. S3 was also consistently less hydrated in complex with the modified DNAs than the corresponding PU.1 complexes. These results prove that by swapping specific amino acids within the DNA binding domain, it is possible to alter the sequence selectivity of the system. S3 does not directly interact with the core consensus sequence 5'-GGAA- '3 but instead contacts the 5' flanking region upstream of the consensus sequence. This demonstrates the potential to adjust protein/DNA complexes through modification of sites distal to the core binding sequence.

4.2 The hydration network of PU.1 is located in the protein/DNA interface in the major groove of the DNA

The work performed with the modified DNA sequences has provided valuable information not only on the role but also the location of the hydration network iconic to PU.1. The different deoxyguanosine derivatives were used to probe direct readout, indirect readout, and direct and indirect readout simultaneously. Modifications were made at the second guanosine of the 5'-GGAA-3'- this guanosine was chosen due to the conserved interaction between arginine

232 and the carbonyl at position 6 on guanosine. The dI modification was used to test the role of indirect readout through the removal of the amino group located in the minor groove of the forward strand of DNA. d2-AP was selected to probe direct readout by removing the direct interaction between the carbonyl on guanosine and Arg 232. Iso-dG is the only construct that alters both direct and indirect readout by swapping the positions of the carbonyl and the amino group on guanosine.

When the dI modified sequence was paired with PU.1, the binding affinity dropped, but the hydration level was mainly left unmodified. This shows that the hydration network of PU.1 is located in the major groove of DNA in the protein/DNA interface rather than the minor groove. Suppose there was a substantial level of hydration located in the minor groove around the consensus sequence. In that case, it would be expected that the loss of the amino group would alter hydration noticeably. The lack of any significant change in hydration in the PU.1/dI complex indicates the minor groove is not implicated in the hydration network. S3 is a consistently less hydrated version of PU.1 but also shows a similar hydration pattern when bound to dI DNA. Both the S3/HA and S3/dI complex retain similar hydration levels, again suggesting that hydration is not in the minor groove and that the limited hydration in the S3 DNA complexes should be in the major groove.

Both PU.1 and S2 with the d2-AP sequence saw significantly decreased hydration and binding affinity. In both cases, the d2-AP/dC complexes showed hydration comparable to the LA complex but a higher binding affinity. When bound to d2-AP/dT, the level of hydration dips below that in the LA complex, but has a higher affinity relative to either the LA or d2-AP/dC complexes. These observations provide valuable insight both into the role of direct readout for binding affinity and the role of direct readout in the hydration network. As S3 is less sequence

selective, these results follow the trend of stronger binding affinities. Slight decreases were seen in the S3/d2-AP/dT complex compared to the S3/d2-AP/dC complex, but still a significant decrease to the unmodified sequence. PU.1 bound to these DNA sequences also displayed a slightly stronger binding affinity when bound to d2-AP/dT than d2-AP/dC, but both binding affinities remain close to the LA complex. These results indicate the direct interaction is critical to high-affinity interactions and losing this essential interaction renders high-affinity binding unviable. The loss of hydration is also interesting, showing that the carbonyl on the second guanosine is vital to forming and retaining the hydration network.

The pairing of the changes to both direct and indirect readout displayed similar results to the loss of direct readout individually. In the PU.1/Iso-dG complex, hydration is decreased to a level matching the d2-AP/dT complex and a binding affinity matching the LA complex. The S3/Iso-dG complex has a binding affinity slightly greater than the S3/LA complex and similar to the S3/d2-AP/dT complex but still showed a complete loss of the hydration network. These results indicate that just retaining canonical Watson-Crick base pairing is not enough to maintain high-affinity interactions or recapture the hydration network lost with the loss of the direct interaction individually.

The hydration network of the PU.1/HA complex is therefore expected to be located in the major groove of the DNA in the protein/DNA interface. The loss of the direct interaction between Arg 232 and guanosine is detrimental to the formation of the hydration network and negatively affects high-affinity binding. The loss of indirect readout at this position through the dI construct results in decreased binding affinity, possibly due to a slight destabilization of the DNA, but is not implicated in the hydration network. Proper readout of the DNA is essential for forming high-affinity complexes and retaining the hydration network in the PU.1/HA complex.

The removal of a single interaction can be enough to cause devastating effects on both hydration and binding affinity.

4.3 Differing thermodynamic patterns characterize high and low-affinity interactions

To gain greater insight into the thermodynamic properties of hydration for sequence selectivity, a series of modified DNA sequences paired with either PU.1 or the S3 chimera, were examined via ITC. ITC was chosen to directly measure changes in enthalpy upon a reaction occurring. In this case, the reaction being studied was the protein interacting with DNA. Performing an ITC experiment to saturation also provided a means of determining the number of binding events and whether the binding events are specific or nonspecific. Once ITC has measured the change in enthalpy (ΔH), it is possible to calculate the change in entropy ($T\Delta S$) of the system using the Gibbs free energy equation, providing a complete thermodynamic analysis of the binding events.

Upon examining the protein and DNA reactions, a pattern emerged. The formation of high-affinity vs. low-affinity complexes had drastically different changes in enthalpy upon forming the 1:1 complex. The high-affinity 1:1 complex is characterized by a negative ΔH or an exothermic reaction in which the system releases energy into the surrounding environment. In contrast, the low-affinity complex is characterized by a positive ΔH or an endothermic reaction in which the system takes up energy from the surrounding environment.

Specific interactions are represented by significant negative changes in Gibbs free energy (ΔG), negative changes in enthalpy (ΔH), and changes in entropy ($T\Delta S$) as low as possible. As specificity decreases, ΔG and ΔH becomes progressively less negative, and $T\Delta S$ becomes more positive. This provides helpful milestones to understand the specificity of the system. The swapping of PU.1 sequence for the corresponding Ets-1 region in the S3 chimera was found to

noticeably decrease sequence specificity and thus had this altered thermodynamic profile compared with PU.1.

4.4 PU.1 and S3 successfully form specific 2:1 complexes on modified DNA sequences

Fluorescence anisotropy experiments performed were calibrated to examine the change in fluorescence upon forming the 1:1 complex on non-labeled DNA. To determine if PU.1 or S3 forms a 2:1 complex on DNA, ITC experiments were performed with DNA held constant, and the protein was titrated into the sample. This method provided a means of distinguishing the 1:1 complex from the 2:1 complex. It was determined that PU.1 and S3 both form the 2:1 complex on modified DNA sequences. This indicates that the 2:1 complex does not require a high-affinity binding site to form since the 2:1 complex is also formed on d2-AP and LA sequences. However, it is not yet clear if S3 bound to d2-AP/dT forms a distinct 2:1 complex, as ΔH for the potential second binding event is very small. More experiments are required to determine if the S3/d2-AP/dT complex does form the 2:1 complex.

The ITC experiments demonstrate the number of protein/DNA complexes that the system formed and whether or not those complexes are specific or nonspecific. This is determined by inflection points in the plot and whether the trend returns to baseline after the complex is reached. If ΔH continues to increase or remains above baseline, this indicates a nonspecific complex. For PU.1, all of the 2:1 complexes were found to be specific complexes. In S3, three of four complexes examined formed specific 2:1 complexes (HA, dI, and LA). The S3/d2-AP/dT complex may be a specific 2:1 complex or a 1:1 complex with a return to baseline. More data points are needed to clarify whether the S3/d2-AP/dT complex is specific or nonspecific. The majority of the combinations examined were found to form specific 2:1 complexes, providing greater insight into the ability of PU.1 and PU.1 chimeras to bind modified DNAs.

Understanding that PU.1 and S3 retain the ability to bind 1:1 followed by 2:1 complexes provides an understanding of the mechanisms and limitations of binding to below optimal DNA sequences.

4.5 Chimeric constructs provide a valuable avenue for examining structurally conserved transcription factors

The use of chimeric constructs in this work proves that it is possible to retain proper secondary structure by replacing the primary sequence, as long as the replacement sequence retains the ability to form the required secondary structure. With this knowledge, it may be possible to design protein constructs containing specific functions and qualities through this chimeric approach. The ETS family retains a highly conserved structure of the DNA binding domain, providing an excellent avenue for testing chimeric models. It is reasonable to believe that this approach could be harnessed in other structurally conserved protein families.

For chimeric constructs to be a viable avenue for examining other trends in other systems, the limitations must be understood. The approach worked in this instance due to the highly conserved structure despite highly divergent primary sequences in the DNA binding domains in the ETS transcription factor family. To apply the same method in another system, a structurally conserved family would need to be considered, which would provide the structural scaffolding required to simply swap sequences. Over the course of this work, chimeras have been designed in which amino acid sequences as small as three residues have been replaced and up to as large as eleven residues. These differing lengths of changes show that various sized swaps are acceptable while still retaining the proper structural registry, at least in the ETS domain scaffolding. Each system could respond differently, and trial and error will be required to determine if the chimeric approach is appropriate for any other given model system in question.

In the case of the effectiveness of the chimeric approach, not all structural elements in the PU.1 DNA binding domain are equally replaceable. The recognition helix (H3) of PU.1 is highly sensitive to changes in the amino acid sequence. While the H3 chimera does retain the proper structure as PU.1, when examined with high-affinity DNA under increased osmotic pressure, it was discovered that the binding affinity had significantly decreased. The binding affinity had been reduced to a level comparable to the low-affinity level of interaction when H3 was exchanged with the corresponding recognition helix of Ets-1. These results show that in PU.1, the recognition helix is the most stringent and that the primary sequence for this specific structure is not random, but rather each residue has an essential role to play. And that by swapping out these specific residues with the corresponding residues from Ets-1, there is not enough overlap to adequately retain the proper binding affinity and level of hydration.

Chimeric constructs provide valuable insight into the role specific structures and amino acid sequences play in protein/DNA interactions. This method may be a highly valuable tool in the future to understand a variety of systems. It is possible to pair chimeric constructs with various environmental changes, not limited to changes in osmotic pressure, which can provide valuable insight into the role of protein/DNA interactions. This work could easily be expanded to determine the role of small molecules on protein/DNA interactions to determine how and where they bind to aid in developing future therapeutics.

4.6 Future Directions

4.6.1 Determining the location and structure of the hydration network via X-ray crystallography

Through the work with chimeric constructs of PU.1, we have determined the secondary structures that control hydration and the role of hydration in the protein-DNA interface in high-

affinity complexes. The bindings, however, did not provide any information as to the formation of the water network, where specifically it is located, and what bases or residues help to form the network. Fluorescence anisotropy experiments have narrowed the location of the hydration network to the major groove of DNA in the protein/DNA interface. The use of modified nucleobases has also shown that the direct interaction between the carbonyl on position 6 of the second guanosine (5'-GGAA-3') is essential for creating and retaining the hydration network of PU.1. This work has also shown that by replacing the S3 structure in PU.1 with the corresponding residues of Ets-1, the hydration network is severely affected, suggesting that S3 plays a vital role in hydration. From this body of evidence, it is understood that the hydration network should be located in the major groove of DNA and extend to the 5'flanking region of DNA.

A method to better study the hydration network in PU.1 is X-ray crystallography. Crystalizing PU.1 and S3 bound to various DNA sequences individually will provide insight into any structural changes. These sequences would include the high-affinity sequence, inosine, and the low-affinity sequence. The PU.1/HA complex offers a baseline for the level and location of hydration. If the PU.1/HA complex is then compared with the PU.1/dI complex, any changes in structure or hydration will result from the difference in the nucleobase, allowing us to understand how the alteration affects hydration and thus binding. The fluorescence anisotropy experiments indicated the change in hydration was minimal, so a significant change in hydration upon forming the complex is not expected in the crystal structure. However, the decrease in binding affinity suggesting either potential unfavorable structural changes that may result in the loss of crucial contacts.

In contrast, comparing PU.1/d2-AP/dT vs. PU.1/HA vs. PU.1/LA complexes would be highly enlightening, as PU.1/d2-AP/dT demonstrates hydration comparable to low-affinity. Comparing these three complexes would provide valuable insight into whether the structure has altered to affect hydration or if the changes in hydration of PU.1/D-2AP/T more closely resemble HA or LA. This would explain the decreased binding affinity, decreased level of hydration, and prove the importance of the direct interaction of the carbonyl on guanosine with Arg 232.

Currently, there are several crystal structures of PU.1 in complex with DNA, but no structures of the chimeric constructs with or without DNA. The lack of structural information on any of the chimeras hinders a better understanding of the role of secondary structural elements have on binding affinity and the pattern of hydration. Moving forward, a crystal structure of S3 bound to high-affinity DNA would be invaluable to understanding the role of S3 in high-affinity interactions and the development of the hydration network of PU.1. By comparing the PU.1/HA complex with the S3/HA complex, it would be possible to establish any structural changes that were not detected by circular dichroism. This comparison will also clarify S3's role in maintaining the hydration network by exploring the differences in water-mediated interactions between PU.1 and Ets-1 residues.

A comparison of the PU.1/dI vs. S3/dI complexes would also be highly enlightening, as the PU.1/dI complex shows decreased binding affinity despite that the alteration is not directly interacting with the recognition helix. However, the S3/dI complex retains high-affinity binding with reduced levels of hydration. By comparing these two crystal structures, it will hopefully be possible to understand how S3 compensates on the flanking region despite the lack of water, while PU.1 is unable to. This would also assist in understanding the variations between Ets-1 and

PU.1 in sequence sensitivity and form a complete picture of the interplay between high-affinity protein/DNA complexes and hydration in PU.1.

By examining these crystal structures in comparison with the PU.1 high-affinity complex, it will be possible to investigate how the hydration network changes between the well-hydrated high-affinity complex, the low hydrated but high-affinity binder inosine complex, and the low hydration and low-affinity binding low-affinity complex. Through the examination and comparison of these three complexes, it is possible to cover the extremes and middle of how PU.1 binds with DNA and provides greater insight into hydration in the protein-DNA interaction.

4.6.2 Is the volume of the protein/DNA complex modified by changes in DNA sequences?

This work has shown how important hydration is in PU.1 for sequence specificity and forming the high-affinity complex with DNA. However, what is not known is the actual volume that the water is taking up within the complex and whether there is a change in volume depending on the complex formed. To explore this volumetric aspect of protein/DNA interaction, the use of a high precision densimeter would provide great insight. Over the course of the experiment, the density of the protein and DNA alone is measured, followed by a measurement of the density of the complex. These densities are then converted to volume as a means to determine the amount of volume taken up by the complex. By examining the proteins and sequences individually, followed by an analysis of the complex, information can be gleaned about how the volume changes between the unbound and bound protein.

With a more complete understanding of how the volume changes upon complex formation, it will then be possible to understand when the hydration network of PU.1 is formed. It is possible that the hydration network is already present around the protein or DNA by itself

and is stabilized upon the formation of the complex. Alternatively, the hydration network may form in parallel to complex formation, and hydrating waters are taken up from bulk solution as a means of stabilizing the protein on the high-affinity DNA. By analyzing the changes in volume between the unbound and bound complex, it is possible to shed light on this mechanism.

Once an understanding of the creation of the hydration network is determined, it will then be possible to analyze the differences between the various proteins and complexes. Fluorescence polarization has indicated an observable difference in the level of hydration between PU.1 and S3. By examining these proteins separately, it will be possible to determine whether one is inherently more hydrated in the unbound state or if they are equally hydrated in the unbound state, and the differences occur upon binding.

Another question that may be answered by high precision densimetry is how drastic the hydration change is between PU.1 and S3 when bound to modified DNA sequences. The fluorescence polarization experiments indicated similar levels of hydration when comparing PU.1 and S3 bound to the same modified sequences. By examining these complexes for changes in volume, it will be possible to determine if these trends hold true under differing experimental conditions. This information regarding the hydration network from volumetric analysis alongside structural analysis by X-ray crystallography will present a more complete picture of the mechanics and driving forces and the specific sites of the hydration network as found in the PU.1/HA complex.

4.6.3 Understanding the dynamic properties of hydration and the change in binding via nuclear magnetic resonance

At present, all of the data collected, and the planned experiments focuses on information obtained at equilibrium and does not provide any insight into the system's dynamics. For this

reason, a valuable next step would be to employ $^{1}H/^{15}N$ heteronuclear NMR spectroscopy to compare the dynamics of the bound and unbound structures. By examining a $^{1}H/^{15}N$ labeled protein, it is possible to determine T_1, T_2, and NOEs of the backbone amides, which by the model-free approach will determine both the magnitude and location of dynamics in the protein [133, 134]. Knowing the dynamics in a reference state (unbound), it is possible to determine how the protein responds upon binding various DNA sequences.

The comparison of unbound PU.1 vs. S3 in solution will provide valuable information on if the dynamics are altered by the S3 chimera relative to WT PU.1, and if so how the inclusion of the Ets-1 residues affects distal regions of the protein. If there is a noticeable change in the dynamics of the S3 chimera, understanding where that change is occurring will provide valuable insight into whether the altered behavior of the S3 construct is due to a localized or overall change to the protein itself. Comparison of PU.1 and S3 bound to HA, dI, d2-AP/dT, and LA vs. the unbound structures, individually and in relation to each other will provide a treasure trove of information on how the varying changes influence each complex. Comparing the $^{1}H/^{15}N$ heteronuclear NMR data vs. the thermodynamic data, it will be possible to better understand how the protein interacts with the DNA, and whether the changes in the DNA sequences are affecting the overall dynamics of the complex or more localized factors.

The thermodynamic analysis of the complexes suggests that hydration and DNA dynamics play an important role in the level of protein/DNA specificity. A comparison of the dynamic data by $^{1}H/^{15}N$ heteronuclear NMR will provide detailed information about any changes in the magnitude, and the location of those dynamics. One of the major questions left from these studies is whether the S3 chimera interacts with the 5’ flanking region of DNA in a different manner, or if an allosteric effect is responsible for the modified binding patterns. The dynamics

data from NMR and the structural data from the crystal structure, along with the change in hydration and thermodynamic profile, will provide a complete picture of how the addition of the Ets-1 residues affect complex formation. Pairing the NMR data with the crystal structure, binding experiments, and volumetric data will provide an extremely detailed, well-rounded view of the mechanism and manner in which PU.1 forms a complex with DNA and how that differs from the complexes formed by S3 with DNA.

4.6.4 Molecular dynamics simulations provide a means of determining the level of DNA dynamics in sequences containing modified deoxyguanosine derivatives

The current understanding of the role of hydration with DNA dynamics comes from inferring that the removal of specific groups on DNA bases results in increased dynamics from other NMR and molecular dynamic (MD) studies. The role of hydration and DNA dynamics has not been directly shown in this system. A means to fill in that missing information is to perform MD simulations on the modified DNA sequences alone and in complex with PU.1 and S3. To perform MD simulations on the modified DNA sequences, each of the modified bases will first have to be parameterized in the MD system of choice, which in this example is Amber. With the parameterization of the modified bases added to the Amber system, it will then be possible to simulate the DNA sequences and to directly measure the time scale and length of base opening and level of DNA dynamics of the modified sequences vs. the high-affinity sequence.

Once the simulations of the individual sequences are performed, it will be possible to examine how the DNA dynamics are altered with either PU.1 or S3. An MD simulation of the protein-DNA complexes will provide greater insight into the level of DNA dynamics and the level and location of hydration in the system. Which will then give an avenue to compare the level and location of hydration from the MD simulation to the crystal structure. In addition, a

comparison of the simulation data and the crystal structure will provide confidence in the MD simulations ability to simulate binding events. By combining the data obtained by MD simulations, crystallography, NMR, volumetrics, fluorescence anisotropy, and ITC, a complete picture of the role and location of hydration and DNA dynamics will be obtained, providing greater understanding and a well-rounded study of the role of hydration and DNA in protein/DNA specificity.

REFERENCES

1. Lin, M. and J.-T. Guo, *New insights into protein-DNA binding specificity from hydrogen bond based comparative study.* Nucleic acids research, 2019. **47**(21): p. 11103-11113.
2. Ahmad, S., H. Kono, M.J. Araúzo-Bravo, and A. Sarai, *ReadOut: structure-based calculation of direct and indirect readout energies and specificities for protein-DNA recognition.* Nucleic acids research, 2006. **34**(Web Server issue): p. W124-W127.
3. Härd, T. and T. Lundbäck, *Thermodynamics of sequence-specific protein-DNA interactions.* Biophysical Chemistry, 1996. **62**(1): p. 121-139.
4. Eaton, B.E., L. Gold, and D.A. Zichi, *Let's get specific: the relationship between specificity and affinity.* Chem Biol, 1995. **2**(10): p. 633-8.
5. Saravanan, M., K. Vasu, and V. Nagaraja, *Evolution of sequence specificity in a restriction endonuclease by a point mutation.* Proceedings of the National Academy of Sciences, 2008. **105**(30): p. 10344-10347.
6. Davis, L.G., M.D. Dibner, and J.F. Battey, *SECTION 5-4 - Restriction Endonucleases (REs) and Their Use*, in *Basic Methods in Molecular Biology*, L.G. Davis, M.D. Dibner, and J.F. Battey, Editors. 1986, Elsevier. p. 51-57.
7. Stephenson, F.H., *9 - Recombinant DNA*, in *Calculations for Molecular Biology and Biotechnology*, F.H. Stephenson, Editor. 2003, Academic Press: Burlington. p. 186-241.
8. Rohs, R., X. Jin, S.M. West, R. Joshi, B. Honig, and R.S. Mann, *Origins of specificity in protein-DNA recognition.* Annual Review of Biochemistry, 2010. **79**: p. 233-269.
9. Bareket-Samish, A., I. Cohen, and T.E. Haran, *Direct versus indirect readout in the interaction of the trp repressor with non-canonical binding sites11Edited by J. Karn.* Journal of Molecular Biology, 1998. **277**(5): p. 1071-1080.
10. Chu, X. and J. Wang, *Specificity and Affinity Quantification of Flexible Recognition from Underlying Energy Landscape Topography.* PLOS Computational Biology, 2014. **10**(8): p. e1003782.
11. Jayaram, B. and T. Jain, *The role of water in protein-DNA recognition.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 343-61.
12. Demidov, V.V. and M.D. Frank-Kamenetskii, *Two sides of the coin: affinity and specificity of nucleic acid interactions.* Trends in Biochemical Sciences, 2004. **29**(2): p. 62-71.
13. Jantz, D. and J.M. Berg, *Probing the DNA-binding affinity and specificity of designed zinc finger proteins.* Biophysical journal, 2010. **98**(5): p. 852-860.
14. Koudelka, G.B., S.A. Mauro, and M. Ciubotaru, *Indirect Readout of DNA Sequence by Proteins: The Roles of DNA Sequence-Dependent Intrinsic and Extrinsic Forces*, in *Progress in Nucleic Acid Research and Molecular Biology*. 2006, Academic Press. p. 143-177.
15. Steffen, N.R., S.D. Murphy, L. Tolleri, G.W. Hatfield, and R.H. Lathrop, *DNA sequence and structure: direct and indirect recognition in protein-DNA binding.* Bioinformatics, 2002. **18 Suppl 1**: p. S22-30.
16. Stormo, G.D., *Modeling the specificity of protein-DNA interactions.* Quantitative biology (Beijing, China), 2013. **1**(2): p. 115-130.
17. Prieto, J., P. Redondo, B. López-Méndez, M. D'Abramo, N. Merino, F.J. Blanco, P. Duchateau, G. Montoya, and R. Molina, *Understanding the indirect DNA read-out specificity of I-CreI Meganuclease.* Scientific Reports, 2018. **8**(1): p. 10286.

18. Yamasaki, S., T. Terada, H. Kono, K. Shimizu, and A. Sarai, *A new method for evaluating the specificity of indirect readout in protein-DNA recognition.* Nucleic Acids Res, 2012. **40**(17): p. e129.
19. Lindemose, S., P.E. Nielsen, and N.E. Møllegaard, *Dissecting direct and indirect readout of cAMP receptor protein DNA binding using an inosine and 2,6-diaminopurine in vitro selection system.* Nucleic Acids Res, 2008. **36**(14): p. 4797-807.
20. Poon, G.M. and R.B. Macgregor, Jr., *Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1.* J Mol Biol, 2003. **328**(4): p. 805-19.
21. Dellarole, M., I.E. Sánchez, and G. de Prat Gay, *Thermodynamics of cooperative DNA recognition at a replication origin and transcription regulatory site.* Biochemistry, 2010. **49**(48): p. 10277-86.
22. Kim, S.S., J.K. Tam, A.F. Wang, and R.S. Hegde, *The structural basis of DNA target discrimination by papillomavirus E2 proteins.* J Biol Chem, 2000. **275**(40): p. 31245-54.
23. Zhang, Y., Z. Xi, R.S. Hegde, Z. Shakked, and D.M. Crothers, *Predicting indirect readout effects in protein-DNA interactions.* Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8337-41.
24. Janin, J., *Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition.* Structure, 1999. **7**(12): p. R277-R279.
25. Huggins, David J., *Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations.* Biophysical Journal, 2015. **108**(4): p. 928-936.
26. Dragan, A.I., C.M. Read, and C. Crane-Robinson, *Enthalpy–entropy compensation: the role of solvation.* European Biophysics Journal, 2017. **46**(4): p. 301-308.
27. Kochoyan, M. and J.L. Leroy, *Hydration and solution structure of nucleic acids.* Curr Opin Struct Biol, 1995. **5**(3): p. 329-33.
28. Yonetani, Y. and H. Kono, *Sequence dependencies of DNA deformability and hydration in the minor groove.* Biophysical journal, 2009. **97**(4): p. 1138-1147.
29. Clore, G.M., A. Bax, J.G. Omichinski, and A.M. Gronenborn, *Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA.* Structure, 1994. **2**(2): p. 89-94.
30. Jacobson, A., W. Leupin, E. Liepinsh, and G. Otting, *Minor Groove Hydration of DNA in Aqueous Solution: Sequence-dependent Next Neighbor Effect of the Hydration Lifetimes in d(TTAA)2 Segments Measured by NMR Spectroscopy.* Nucleic Acids Research, 1996. **24**(15): p. 2911-2918.
31. Woda, J., B. Schneider, K. Patel, K. Mistry, and H.M. Berman, *An Analysis of the Relationship between Hydration and Protein-DNA Interactions.* Biophysical Journal, 1998. **75**(5): p. 2170-2177.
32. Poon, G.M.K., P. Gross, and R.B. Macgregor, *The Sequence-Specific Association of the ETS Domain of Murine PU.1 with DNA Exhibits Unusual Energetics.* Biochemistry, 2002. **41**(7): p. 2361-2371.
33. Rand*, R.P., V.A. Parsegian, and D.C. Rau, *Intracellular osmotic action.* Cellular and Molecular Life Sciences CMLS, 2000. **57**(7): p. 1018-1032.
34. Fried, M.G., D.F. Stickle, K.V. Smirnakis, C. Adams, D. MacDonald, and P. Lu, *Role of hydration in the binding of lac repressor to DNA.* J Biol Chem, 2002. **277**(52): p. 50676-82.

35. Parsegian, V.A., R.P. Rand, and D.C. Rau, *Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives.* Proceedings of the National Academy of Sciences, 2000. **97**(8): p. 3987-3992.
36. Robinson, C.R. and S.G. Sligar, *Changes in solvation during DNA binding and cleavage are critical to altered specificity of the EcoRI endonuclease.* Proceedings of the National Academy of Sciences, 1998. **95**(5): p. 2186-2191.
37. Robinson, C.R. and S.G. Sligar, *Hydrostatic Pressure Reverses Osmotic Pressure Effects on the Specificity of EcoRI-DNA Interactions.* Biochemistry, 1994. **33**(13): p. 3787-3793.
38. Sidorova, N.Y. and D.C. Rau, *Differences in water release for the binding of EcoRI to specific and nonspecific DNA sequences.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(22): p. 12272-12277.
39. Sidorova, N.Y., S. Muradymov, and D.C. Rau, *Differences in hydration coupled to specific and nonspecific competitive binding and to specific DNA Binding of the restriction endonuclease BamHI.* J Biol Chem, 2006. **281**(47): p. 35656-66.
40. Sidorova, N.Y. and D.C. Rau, *Differences between EcoRI Nonspecific and "Star" Sequence Complexes Revealed by Osmotic Stress.* Biophysical Journal, 2004. **87**(4): p. 2564-2576.
41. Brown, M.P., A.O. Grillo, M. Boyer, and C.A. Royer, *Probing the role of water in the tryptophan repressor-operator complex.* Protein science : a publication of the Protein Society, 1999. **8**(6): p. 1276-1285.
42. Rhodes, M.M., K. Réblová, J. Šponer, and N.G. Walter, *Trapped water molecules are essential to structural dynamics and function of a ribozyme.* Proceedings of the National Academy of Sciences, 2006. **103**(36): p. 13380-13385.
43. !!! INVALID CITATION !!! [3, 33].
44. Kumar, R., *Role of naturally occurring osmolytes in protein folding and stability.* Archives of Biochemistry and Biophysics, 2009. **491**(1): p. 1-6.
45. Timasheff, S.N., *Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components.* Proceedings of the National Academy of Sciences, 2002. **99**(15): p. 9721-9726.
46. Chalikian, T.V., A.P. Sarvazyan, G.E. Plum, and K.J. Breslauer, *Influence of base composition, base sequence, and duplex structure on DNA hydration: apparent molar volumes and apparent molar adiabatic compressibilities of synthetic and natural DNA duplexes at 25 degrees C.* Biochemistry, 1994. **33**(9): p. 2394-401.
47. Chalikian, T.V. and R.B. Macgregor, *Nucleic acid hydration: a volumetric perspective.* Physics of Life Reviews, 2007. **4**(2): p. 91-115.
48. Toleikis, Z., V.A. Sirotkin, G. Skvarnavičius, J. Smirnovienė, C. Roumestand, D. Matulis, and V. Petrauskas, *Volume of Hsp90 Protein-Ligand Binding Determined by Fluorescent Pressure Shift Assay, Densitometry, and NMR.* J Phys Chem B, 2016. **120**(37): p. 9903-12.
49. Chalikian, T.V., *Volumetric Properties of Proteins.* Annual Review of Biophysics and Biomolecular Structure, 2003. **32**(1): p. 207-235.
50. Mitra, L., J.B. Rouget, B. Garcia-Moreno, C.A. Royer, and R. Winter, *Towards a quantitative understanding of protein hydration and volumetric properties.* Chemphyschem, 2008. **9**(18): p. 2715-21.

51. Soto, A.M., B.I. Kankia, P. Dande, B. Gold, and L.A. Marky, *Thermodynamic and hydration effects for the incorporation of a cationic 3-aminopropyl chain into DNA.* Nucleic Acids Res, 2002. **30**(14): p. 3171-80.
52. Calero-Rubio, C., C. Strab, G.V. Barnett, and C.J. Roberts, *Protein Partial Molar Volumes in Multicomponent Solutions from the Perspective of Inverse Kirkwood–Buff Theory.* The Journal of Physical Chemistry B, 2017. **121**(24): p. 5897-5907.
53. Chalikian, T.V., J. Völker, A.R. Srinivasan, W.K. Olson, and K.J. Breslauer, *The hydration of nucleic acid duplexes as assessed by a combination of volumetric and structural techniques.* Biopolymers, 1999. **50**(5): p. 459-71.
54. Otwinowski, Z., R.W. Schevitz, R.G. Zhang, C.L. Lawson, A. Joachimiak, R.Q. Marmorstein, B.F. Luisi, and P.B. Sigler, *Crystal structure of trp repressor/operator complex at atomic resolution.* Nature, 1988. **335**(6188): p. 321-9.
55. Ha, V.L.T., N. Erlitzki, A.A. Farahat, A. Kumar, D.W. Boykin, and G.M.K. Poon, *Dissecting Dynamic and Hydration Contributions to Sequence-Dependent DNA Minor Groove Recognition.* Biophysical Journal, 2020. **119**(7): p. 1402-1415.
56. Mota, P.C., O. Ferreira, L. Hnědkovský, S.P. Pinho, and I. Cibulka, *Partial Molar Volumes of l-Serine and l-Threonine in Aqueous Ammonium Sulfate Solutions at (278.15, 288.15, 298.15, and 308.15) K.* Journal of Solution Chemistry, 2014. **43**(2): p. 283-297.
57. Ladbury, J.E. and B.Z. Chowdhry, *Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions.* Chem Biol, 1996. **3**(10): p. 791-801.
58. Poon, G.M., *DNA binding regulates the self-association of the ETS domain of PU.1 in a sequence-dependent manner.* Biochemistry, 2012. **51**(20): p. 4096-107.
59. Kozlov, A.G. and T.M. Lohman, *SSB binding to ssDNA using isothermal titration calorimetry.* Methods in molecular biology (Clifton, N.J.), 2012. **922**: p. 37-54.
60. Fodor, E. and A. Ginsburg, *Specific DNA binding by the homeodomain Nkx2.5(C56S): detection of impaired DNA or unfolded protein by isothermal titration calorimetry.* Proteins, 2006. **64**(1): p. 13-8.
61. Morton, C.J. and J.E. Ladbury, *Water-mediated protein-DNA interactions: the relationship of thermodynamics to structural detail.* Protein science : a publication of the Protein Society, 1996. **5**(10): p. 2115-2118.
62. Nakayama, N., G. Sakashita, T. Nagata, N. Kobayashi, H. Yoshida, S.Y. Park, Y. Nariai, H. Kato, E. Obayashi, K. Nakayama, S. Kyo, and T. Urano, *Nucleus Accumbens-Associated Protein 1 Binds DNA Directly through the BEN Domain in a Sequence-Specific Manner.* Biomedicines, 2020. **8**(12).
63. Degnan, B.M., S.M. Degnan, T. Naganuma, and D.E. Morse, *The ets multigene family is conserved throughout the Metazoa.* Nucleic Acids Research, 1993. **21**(15): p. 3479-3484.
64. Poon, G.M.K. and H.M. Kim, *Signatures of DNA target selectivity by ETS transcription factors.* Transcription, 2017. **8**(3): p. 193-203.
65. Esaki, S., M.G. Evich, N. Erlitzki, M.W. Germann, and G.M.K. Poon, *Multiple DNA-binding modes for the ETS family transcription factor PU.1.* J Biol Chem, 2017. **292**(39): p. 16044-16054.
66. Wang, S., M.H. Linde, M. Munde, V.D. Carvalho, W.D. Wilson, and G.M. Poon, *Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1.* J Biol Chem, 2014. **289**(31): p. 21605-16.

67. Hollenhorst, P.C., L.P. McIntosh, and B.J. Graves, *Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors.* Annual Review of Biochemistry, 2011. **80**(1): p. 437-471.
68. Hollenhorst, P.C., K.J. Chandler, R.L. Poulsen, W.E. Johnson, N.A. Speck, and B.J. Graves, *DNA specificity determinants associate with distinct transcription factor functions.* PLoS Genet, 2009. **5**(12): p. e1000778.
69. Szymczyna, B.R. and C.H. Arrowsmith, *DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition.* J Biol Chem, 2000. **275**(37): p. 28363-70.
70. Wei, G.H., G. Badis, M.F. Berger, T. Kivioja, K. Palin, M. Enge, M. Bonke, A. Jolma, M. Varjosalo, A.R. Gehrke, J. Yan, S. Talukder, M. Turunen, M. Taipale, H.G. Stunnenberg, E. Ukkonen, T.R. Hughes, M.L. Bulyk, and J. Taipale, *Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo.* Embo j, 2010. **29**(13): p. 2147-60.
71. Garvie, C.W., J. Hagman, and C. Wolberger, *Structural studies of Ets-1/Pax5 complex formation on DNA.* Mol Cell, 2001. **8**(6): p. 1267-76.
72. Xhani, S., S. Esaki, K. Huang, N. Erlitzki, and G.M. Poon, *Distinct Roles for Interfacial Hydration in Site-Specific DNA Recognition by ETS-Family Transcription Factors.* J Phys Chem B, 2017. **121**(13): p. 2748-2758.
73. He, G., A. Tolic, J.K. Bashkin, and Gregory M K. Poon, *Heterogeneous dynamics in DNA site discrimination by the structurally homologous DNA-binding domains of ETS-family transcription factors.* Nucleic Acids Research, 2015. **43**(8): p. 4322-4331.
74. Hollenhorst, P.C., A.A. Shah, C. Hopkins, and B.J. Graves, *Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family.* Genes & development, 2007. **21**(15): p. 1882-1894.
75. Carrère, S., A. Verger, A. Flourens, D. Stehelin, and M. Duterque-Coquillaud, *Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes via two distinct domains.* Oncogene, 1998. **16**(25): p. 3261-8.
76. Nye, J.A., J.M. Petersen, C.V. Gunther, M.D. Jonsen, and B.J. Graves, *Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif.* Genes Dev, 1992. **6**(6): p. 975-90.
77. Kodandapani, R., F. Pio, C.Z. Ni, G. Piccialli, M. Klemsz, S. McKercher, R.A. Maki, and K.R. Ely, *A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex.* Nature, 1996. **380**(6573): p. 456-60.
78. Munde, M., S. Wang, A. Kumar, C.E. Stephens, A.A. Farahat, D.W. Boykin, W.D. Wilson, and G.M. Poon, *Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines.* Nucleic Acids Res, 2014. **42**(2): p. 1379-90.
79. Stephens, D.C. and G.M. Poon, *Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains.* Nucleic Acids Res, 2016. **44**(18): p. 8671-8681.
80. Kang, H.-S., M.L. Nelson, C.D. Mackereth, M. Schärpf, B.J. Graves, and L.P. McIntosh, *Identification and structural characterization of a CBP/p300-binding domain from the ETS family transcription factor GABP alpha.* Journal of molecular biology, 2008. **377**(3): p. 636-646.

81. Mackereth, C.D., M. Schärpf, L.N. Gentile, S.E. MacIntosh, C.M. Slupsky, and L.P. McIntosh, *Diversity in structure and function of the Ets family PNT domains.* J Mol Biol, 2004. **342**(4): p. 1249-64.
82. Fisher, R.C. and E.W. Scott, *Role of PU.1 in hematopoiesis.* Stem Cells, 1998. **16**(1): p. 25-37.
83. Antony-Debré, I., A. Paul, J. Leite, K. Mitchell, H.M. Kim, L.A. Carvajal, T.I. Todorova, K. Huang, A. Kumar, A.A. Farahat, B. Bartholdy, S.R. Narayanagari, J. Chen, A. Ambesi-Impiombato, A.A. Ferrando, I. Mantzaris, E. Gavathiotis, A. Verma, B. Will, D.W. Boykin, W.D. Wilson, G.M. Poon, and U. Steidl, *Pharmacological inhibition of the transcription factor PU.1 in leukemia.* J Clin Invest, 2017. **127**(12): p. 4297-4313.
84. Poon, G.M. and R.B. Macgregor, Jr., *A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1.* J Mol Biol, 2004. **335**(1): p. 113-27.
85. Poon, G.M.K., *Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface.* The Journal of biological chemistry, 2012. **287**(22): p. 18297-18307.
86. Mohaghegh, N., D. Bray, J. Keenan, A. Penvose, K.K. Andrilenas, V. Ramlall, and T. Siggers, *Biophysical Principles of Lineage Factor PU.1 Binding Revealed by NextPBMs.* bioRxiv, 2018: p. 328625.
87. Pham, T.H., J. Minderjahn, C. Schmidl, H. Hoffmeister, S. Schmidhofer, W. Chen, G. Längst, C. Benner, and M. Rehli, *Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1.* Nucleic Acids Res, 2013. **41**(13): p. 6391-402.
88. Xhani, S., S. Lee, H.M. Kim, S. Wang, S. Esaki, V.L.T. Ha, M. Khanezarrin, G.L. Fernandez, A.V. Albrecht, J.M. Aramini, M.W. Germann, and G.M.K. Poon, *Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1.* Science Advances, 2020. **6**(8): p. eaay3178.
89. Sidorova, N.Y., S. Muradymov, and D.C. Rau, *Solution parameters modulating DNA binding specificity of the restriction endonuclease EcoRV.* Febs j, 2011. **278**(15): p. 2713-27.
90. Robinson, C.R. and S.G. Sligar, *Heterogeneity in molecular recognition by restriction endonucleases: osmotic and hydrostatic pressure effects on BamHI, Pvu II, and EcoRV specificity.* Proceedings of the National Academy of Sciences, 1995. **92**(8): p. 3444-3448.
91. Lynch, T.W. and S.G. Sligar, *Macromolecular hydration changes associated with BamHI binding and catalysis.* J Biol Chem, 2000. **275**(39): p. 30561-5.
92. Shakked, Z., G. Guzikevich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P.B. Sigler, *Determinants of repressor/operator recognition from the structure of the trp operator binding site.* Nature, 1994. **368**(6470): p. 469-73.
93. Whyte, Warren A., David A. Orlando, D. Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong I. Lee, and Richard A. Young, *Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.* Cell, 2013. **153**(2): p. 307-319.
94. Weigelt, K., M. Lichtinger, M. Rehli, and T. Langmann, *Transcriptomic profiling identifies a PU.1 regulatory network in macrophages.* Biochemical and Biophysical Research Communications, 2009. **380**(2): p. 308-312.

95. Garrett-Sinha, L.A., G.H. Su, S. Rao, S. Kabak, Z. Hao, M.R. Clark, and M.C. Simon, *PU.1 and Spi-B are required for normal B cell receptor-mediated signal transduction.* Immunity, 1999. **10**(4): p. 399-408.
96. Rao, S., A. Matsumura, J. Yoon, and M.C. Simon, *SPI-B activates transcription via a unique proline, serine, and threonine domain and exhibits DNA binding affinity differences from PU.1.* J Biol Chem, 1999. **274**(16): p. 11115-24.
97. Findlay, V.J., A.C. LaRue, D.P. Turner, P.M. Watson, and D.K. Watson, *Understanding the role of ETS-mediated gene regulation in complex biological processes.* Adv Cancer Res, 2013. **119**: p. 1-61.
98. Anderson, M.K., G. Hernandez-Hoyos, R.A. Diamond, and E.V. Rothenberg, *Precise developmental regulation of Ets family transcription factors during specification and commitment to the T cell lineage.* Development, 1999. **126**(14): p. 3131-48.
99. Ciau-Uitz, A., L. Wang, R. Patient, and F. Liu, *ETS transcription factors in hematopoietic stem cell development.* Blood Cells Mol Dis, 2013. **51**(4): p. 248-55.
100. He, G., A. Tolic, J.K. Bashkin, and G.M. Poon, *Heterogeneous dynamics in DNA site discrimination by the structurally homologous DNA-binding domains of ETS-family transcription factors.* Nucleic Acids Res, 2015. **43**(8): p. 4322-31.
101. Ho, S.N., *Intracellular water homeostasis and the mammalian cellular osmotic stress response.* J Cell Physiol, 2006. **206**(1): p. 9-15.
102. Go, W.Y., X. Liu, M.A. Roti, F. Liu, and S.N. Ho, *NFAT5/TonEBP mutant mice define osmotic stress as a critical feature of the lymphoid microenvironment.* Proc Natl Acad Sci U S A, 2004. **101**(29): p. 10673-8.
103. Lima, A.F., G. May, J. Díaz-Colunga, S. Pedreiro, A. Paiva, L. Ferreira, T. Enver, F.J. Iborra, and R. Pires das Neves, *Osmotic modulation of chromatin impacts on efficiency and kinetics of cell fate modulation.* Scientific Reports, 2018. **8**(1): p. 7210.
104. Stephens, D.C., H.M. Kim, A. Kumar, A.A. Farahat, D.W. Boykin, and G.M. K. Poon, *Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis.* Nucleic Acids Research, 2016. **44**(9): p. 4005-4013.
105. Stephens, D.C., H.M. Kim, A. Kumar, A.A. Farahat, D.W. Boykin, and G.M.K. Poon, *Pharmacologic efficacy of PU.1 inhibition by heterocyclic dications: a mechanistic analysis.* Nucleic Acids Res, 2016. **44**(9): p. 4005-13.
106. Poon, G.M., *Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface.* J Biol Chem, 2012. **287**(22): p. 18297-307.
107. Yung-Chi, C. and W.H. Prusoff, *Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction.* Biochem Pharmacol, 1973. **22**(23): p. 3099-3108.
108. Wells, J.W., *Analysis and interpretation of binding at equilibrium*, in *Receptor-Ligand Interactions: a Practical Approach*, E.C. Hulme, Editor. 1992, IRL Press at Oxford University Press: Oxford [England]; New York. p. 289-395.
109. Ivani, I., P.D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J.L. Gelpí, C. González, M. Vendruscolo, C.A. Laughton, S.A. Harris, D.A. Case, and M. Orozco, *Parmbsc1: a refined force field for DNA simulations.* Nature Methods, 2016. **13**(1): p. 55-58.

110. Bordoli, L., F. Kiefer, K. Arnold, P. Benkert, J. Battey, and T. Schwede, *Protein structure homology modeling using SWISS-MODEL workspace.* Nat Protoc, 2009. **4**(1): p. 1-13.
111. Weirauch, M.T., A. Yang, M. Albu, A.G. Cote, A. Montenegro-Montero, P. Drewe, H.S. Najafabadi, S.A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.C. Lozano, M. Galli, M.G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J.S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A.J.M. Walhout, F.Y. Bouget, G. Ratsch, L.F. Larrondo, J.R. Ecker, and T.R. Hughes, *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. **158**(6): p. 1431-1443.
112. Workman, C.T., Y. Yin, D.L. Corcoran, T. Ideker, G.D. Stormo, and P.V. Benos, *enoLOGOS: a versatile web tool for energy normalized sequence logos.* Nucleic Acids Res, 2005. **33**(suppl 2): p. W389-W392.
113. Poon, G.M., P. Groß, and R.B. Macgregor, Jr., *The sequence-specific association of the ETS domain of murine PU.1 with DNA exhibits unusual energetics.* Biochemistry, 2002. **41**(7): p. 2361-71.
114. Capp, M.W., L.M. Pegram, R.M. Saecker, M. Kratz, D. Riccardi, T. Wendorff, J.G. Cannon, and M.T. Record, *Interactions of the Osmolyte Glycine Betaine with Molecular Surfaces in Water: Thermodynamics, Structural Interpretation, and Prediction of m-Values.* Biochemistry, 2009. **48**(43): p. 10372-10379.
115. De, S., A.C. Chan, H.J. Coyne, 3rd, N. Bhachech, U. Hermsdorf, M. Okon, M.E. Murphy, B.J. Graves, and L.P. McIntosh, *Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6.* J Mol Biol, 2014. **426**(7): p. 1390-406.
116. Wang, Y., L. Feng, M. Said, S. Balderman, Z. Fayazi, Y. Liu, D. Ghosh, and A.M. Gulick, *Analysis of the 2.0 Å Crystal Structure of the Protein−DNA Complex of the Human PDEF Ets Domain Bound to the Prostate Specific Antigen Regulatory Site.* Biochemistry, 2005. **44**(19): p. 7095-7106.
117. Wang, Y., L. Feng, M. Said, S. Balderman, Z. Fayazi, Y. Liu, D. Ghosh, and A.M. Gulick, *Analysis of the 2.0 A crystal structure of the protein-DNA complex of the human PDEF Ets domain bound to the prostate specific antigen regulatory site.* Biochemistry, 2005. **44**(19): p. 7095-106.
118. Schneider, T.D., G.D. Stormo, L. Gold, and A. Ehrenfeucht, *Information content of binding sites on nucleotide sequences.* J Mol Biol, 1986. **188**(3): p. 415-31.
119. Hertz, G.Z. and G.D. Stormo, *Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.* Bioinformatics, 1999. **15**(7-8): p. 563-77.
120. Jolma, A., J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale, *DNA-binding specificities of human transcription factors.* Cell, 2013. **152**(1-2): p. 327-39.
121. Pio, F., N. Assa-Munt, J. Yguerabide, and R.A. Maki, *Mutants of ETS domain PU.1 and GGAA/T recognition: free energies and kinetics.* Protein science : a publication of the Protein Society, 1999. **8**(10): p. 2098-2109.
122. Monera, O.D., T.J. Sereda, N.E. Zhou, C.M. Kay, and R.S. Hodges, *Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix.* J Pept Sci, 1995. **1**(5): p. 319-29.

123. Escalante, C.R., A.L. Brass, J.M. Pongubala, E. Shatova, L. Shen, H. Singh, and A.K. Aggarwal, *Crystal structure of PU.1/IRF-4/DNA ternary complex.* Mol Cell, 2002. **10**(5): p. 1097-105.
124. Makarov, V., B.M. Pettitt, and M. Feig, *Solvation and Hydration of Proteins and Nucleic Acids: A Theoretical View of Simulation and Experiment.* Accounts of Chemical Research, 2002. **35**(6): p. 376-384.
125. Spolar, R.S. and M.T. Record, Jr., *Coupling of local folding to site-specific binding of proteins to DNA.* Science, 1994. **263**(5148): p. 777-84.
126. Privalov, P.L. and C. Crane-Robinson, *Role of water in the formation of macromolecular structures.* European Biophysics Journal, 2017. **46**(3): p. 203-224.
127. Fuxreiter, M., M. Mezei, I. Simon, and R. Osman, *Interfacial water as a "hydration fingerprint" in the noncognate complex of BamHI.* Biophysical journal, 2005. **89**(2): p. 903-911.
128. Albrecht, A.V., H. Mi Kim, and G.M.K. Poon, *Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach.* Nucleic Acids Res, 2018.
129. Reha-Krantz, L.J., C. Hariharan, U. Subuddhi, S. Xia, C. Zhao, J. Beckman, T. Christian, and W. Konigsberg, *Structure of the 2-aminopurine-cytosine base pair formed in the polymerase active site of the RB69 Y567A-DNA polymerase.* Biochemistry, 2011. **50**(46): p. 10136-10149.
130. Goodman, M.F. and R.L. Ratliff, *Evidence of 2-aminopurine-cytosine base mispairs involving two hydrogen bonds.* Journal of Biological Chemistry, 1983. **258**(21): p. 12842-6.
131. Fagan, P.A., C. Fàbrega, R. Eritja, M.F. Goodman, and D.E. Wemmer, *NMR Study of the Conformation of the 2-Aminopurine:Cytosine Mismatch in DNA.* Biochemistry, 1996. **35**(13): p. 4026-4033.
132. Ferris, Z.E., Q. Li, and M.W. Germann, *Substituting Inosine for Guanosine in DNA: Structural and Dynamic Consequences.* Natural Product Communications, 2019. **14**(5): p. 1934578X19850032.
133. Kharchenko, V., M. Nowakowski, M. Jaremko, A. Ejchart, and Ł. Jaremko, *Dynamic 15N{1H} NOE measurements: a tool for studying protein dynamics.* Journal of Biomolecular NMR, 2020. **74**(12): p. 707-716.
134. Metcalfe, E.E., J. Zamoon, D.D. Thomas, and G. Veglia, *(1)H/(15)N heteronuclear NMR spectroscopy shows four dynamic domains for phospholamban reconstituted in dodecylphosphocholine micelles.* Biophysical journal, 2004. **87**(2): p. 1205-1214.